Princess Sumaya University for Technology
جامعة الأميرة سميّة للتكنولوجيا

صندوق دعم البحث العلمي والابتكار
Scientific Research and Innovation Support Fund

Jordanian Journal of Computers and Information Technology

March 2026 VOLUME 12 NUMBER 01
ISSN 2415 - 1076 (Online)
ISSN 2413 - 9351 (Print)

PAGES **PAPERS**

1 - 17 LOG-STAT: AN ILLUMINATION-BALANCING ALGORITHM FOR ARCHAEOLOGICAL IMAGES CAPTURED IN NON-IDEAL LIGHTING
Zohair Al-Ameen and Basim Mahmood

18 - 30 MULTI-CLASS HEART DISEASE CLASSIFICATION USING MULTI-LEAD ECG FEATURES AND ENSEMBLE LEARNING
Bala V. Isunuri, Venkata S. Alapati, Sri C. Marripudi, Gopala K. Parimi and Akshaya V. Koganti

31 - 51 GEA-COPE: AN EFFECTIVE MODEL FOR CROSS- DOMAIN GRAPH PRE-TRAINING
Yiming Zhao and Yongqing Wu

52 - 66 VEGA-360: VIEWPORT-AWARE HIERARCHICAL GROUPED ALLOCATION FOR MULTI-LAYER 360° VIDEO STREAMING
Phi Dinh Huynh and Nguyen Viet Hung

67 - 81 A SCALABLE FEDERATED DEEP REINFORCEMENT LEARNING ARCHITECTURE FOR COLLABORATIVE LEARNING
Tarek Amine Haddad

82 - 97 AMBIENT BACKSCATTER-ASSISTED PASSIVE RELAYING WITH ENERGY HARVESTING: PERFORMANCE ANALYSIS
Lam-Dong Huynh, Lam-Thanh Tu, Quang-Sang Nguyen and Tan N. Nguyen

98 - 111 BRAIN-TUMOR CLASSIFICATION USING RESNET50 ENHANCED WITH SE AND CBAM ATTENTION MECHANISMS
Nadia S. Abdulsattar and Fatimah S. Abdulsattar

112 - 134 INTERPRETABLE INTRUSION DETECTION WITH TABNET ATTENTION MASKS ENHANCED BY INFORMATION GAIN AND GREY WOLF OPTIMIZATION
Mohamed Goismi, Mohamed Debbab, Moustafa Maaskri and Djamel Seghier

www.jjcit.org jjcit@psut.edu.jo

An International Peer-Reviewed Scientific Journal Financed by the Scientific Research and Innovation Support Fund

Jordanian Journal of Computers and Information Technology (JJCIT)

The Jordanian Journal of Computers and Information Technology (JJCIT) is an international journal that publishes original, high-quality and cutting edge research papers on all aspects and technologies in ICT fields.
JJCIT is hosted and published by Princess Sumaya University for Technology (PSUT) and supported by the Scientific Research Support Fund in Jordan. Researchers have the right to read, print, distribute, search, download, copy or link to the full text of articles. JJCIT permits reproduction as long as the source is acknowledged.

AIMS AND SCOPE

The JJCIT aims to publish the most current developments in the form of original articles as well as review articles in all areas of Telecommunications, Computer Engineering and Information Technology and make them available to researchers worldwide. The JJCIT focuses on topics including, but not limited to: Computer Engineering & Communication Networks, Computer Science & Information Systems and Information Technology and Applications.

INDEXING

JJCIT is indexed in:

ELSEVIER Scopus
INDEXED IN EMERGING SOURCES CITATION (Web of Science) CLARIVATE ANALYTICS
DOAJ DIRECTORY OF OPEN ACCESS JOURNALS
crossref Text and Data Mining
OCLC WorldCat
BASE
ACADEMIA
Google scholar

EDITORIAL BOARD SUPPORT TEAM

LANGUAGE EDITOR	EDITORIAL BOARD SECRETARY
Haydar Al-Momani	Eyad Al-Kouz

All articles in this issue are open access articles distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).

JJCIT ADDRESS

WEBSITE: www.jjcit.org
EMAIL: jjcit@psut.edu.jo
ADDRESS: Princess Sumaya University for Technology, Khalil Saket Street, Al-Jubaiha
B.O. BOX: 1438 Amman 11941 Jordan
TELEPHONE: +962-6-5359949
FAX: +962-6-7295534

EDITORIAL BOARD

Wejdan Abu Elhaija (EIC)
Ahmad Hiasat (Senior Editor)
Aboul Ella Hassanien
Adil Alpkoçak
Adnan Gutub
Adnan Shaout
Christian Boitet
Gian Carlo Cardarilli
Omer Rana
Mohammad Azzeh
Omar A. Alzubi
Ahmed Al-Taani
Lutfi Al-Sharif
Omar S. Al-Kadi
Raed A. Shatnawi
João L. M. P. Monteiro
Leonel Sousa
Omar Al-Jarrah
Angelo Lorusso

INTERNATIONAL ADVISORY BOARD

Ahmed Yassin Al-Dubai
UK

Albert Y. Zomaya
AUSTRALIA

Chip Hong Chang
SINGAPORE

Izzat Darwazeh
UK

Dia Abu Al Nadi
JORDAN

George Ghinea
UK

Hoda Abdel-Aty Zohdy
USA

Saleh Oqeili
JORDAN

João Barroso
PORTUGAL

Karem Sakallah
USA

Khaled Assaleh
UAE

Laurent-Stephane Didier
FRANCE

Lewis Mackenzies
UK

Zoubir Hamici
JORDAN

Korhan Cengiz
TURKEY

Marco Winzker
GERMANY

Marwan M. Krunz
USA

Mohammad Belal Al Zoubi
JORDAN

Michael Ullman
USA

Ali Shatnawi
JORDAN

Mohammed Benaissa
UK

Basel Mahafzah
JORDAN

Nadim Obaid
JORDAN

Nazim Madhavji
CANADA

Ahmad Al Shamali
JORDAN

Othman Khalifa
MALAYSIA

Shahrul Azman Mohd Noah
MALAYSIA

Shambhu J. Upadhyaya
USA

“Opinions or views expressed in papers published in this journal are those of the author(s) and do not necessarily reflect those of the Editorial Board, the host university or the policy of the Scientific Research Support Fund”.

"ما ورد في هذه المجلة يعبر عن آراء الباحثين ولا يعكس بالضرورة آراء هيئة التحرير أو الجامعة أو سياسة صندوق دعم البحث العلمي والابتكار".

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 12, No. 01, March 2026.

LOG-STAT: AN ILLUMINATION-BALANCING ALGORITHM FOR ARCHAEOLOGICAL IMAGES CAPTURED IN NON-IDEAL LIGHTING

Zohair Al-Ameen[1] and Basim Mahmood[2]

(Received: 18-Oct.-2025, Revised: 6-Dec.-2025, Accepted: 23-Dec.-2025)

ABSTRACT

Archaeological images are occasionally captured in environments with non-ideal lighting. This results in imbalanced illumination and a loss of detail. These problems hinder precise operations, such as analysis, interpretation, representation, and 3D modeling. This study introduces a non-complex illumination balancing algorithm called Log-Stat, leveraging logarithmic approaches and statistical methods. It also includes two main phases, one for illumination balancing and the other for tonality adjustment. The first phase utilizes six mathematical equations, while the second phase utilizes four equations, processing the V channel of the image in the HSV color model. Various images have been used to test the algorithm, and a comparison with ten prominent algorithms is achieved, evaluating the outcomes using six measures. The results have shown the success of Log-Stat in different aspects, including fidelity recovery and illumination balance. This allowed better visualization of details, which the unbalanced illumination effect hindered. Integrating appropriate methods and fine-tuning the parameters enabled the Log-Stat to perform in dissimilar illumination situations.

KEYWORDS

Archeology, Image enhancement, Log-Stat, Illumination balancing, Non-ideal lighting.

1. INTRODUCTION

Digital images are an essential tool in archaeology. They can be utilized for documenting, modeling, or recording findings [1]. However, their quality is often compromised, as various degradations may be included. One degradation of interest is the inconsistent illumination. It makes the image appear underlit (too dark) in certain regions, while appearing overlit (too bright) in the other areas [2]. This yields an undesirable appearance and loss of central details, such as textures, affecting the overall visibility [3]. Other image aspects, such as colors, are also affected due to uneven illumination. Colors appear mispresented, leading to struggles in distinguishing and perceiving the actual information precisely [4]. Overall, inconsistent illumination influences the accurate analysis, extraction, and representation of archaeological subjects. As imaging technology advances, the ability to enhance images has become considerable in archaeological research [5]. Digital images enable non-invasive inspection of scenes. This avoids further deterioration and protects integrity. Thus, high-quality images become a key element [6]. This study addresses the illumination balancing issue by exploring logarithmic and statistical methods to improve accuracy and visibility. The lighting situations in archaeological settings are usually far from ideal. This is due to various factors. Natural light can be limited or insufficient, especially in indoor settings. The nature of the photographed object can also affect the lighting conditions. Reflective or shiny surfaces cause glare, whereas textured surfaces cause shadows [7]. Likewise, artificial lighting, such as flashlights, can lead to inconsistent illumination [8].

Uneven illumination poses significant challenges to archaeologists, distorting the colors and obscuring the details [9]. Without adequate balance, the accuracy of image-based analysis is compromised. Restoring images must be carried out effectively to obtain improved-quality images without generating processing errors [10]. Hence, the key contribution of this research is to develop a non-complex algorithm that rapidly and adequately balances the illumination of archaeological images while avoiding limitations, such as shadows around edges, brightness amplification (i.e., global increase in brightness), over-enhancements, and distortions. In this context, balancing the illumination offers the following benefits: (i) it improves clarity and visibility, leading to more reliable interpretations; (ii) it helps reduce

---

1. Z. Al-Ameen is with the ICT Research Unit, Computer Center, University of Mosul Presidency, University of Mosul, Mosul, Nineveh, Iraq. Email: `qizohair@uomosul.edu.iq`

2. B. Mahmood is with the Computer Science Department, College of Computer Science and Mathematics, University of Mosul, Mosul, Nineveh, Iraq. Email: `bmahmood@uomosul.edu.iq`

distortions caused by glare, shadows, and over- or under-exposure, leading to a more realistic appearance. Thus, the Log-Stat algorithm is developed, combining logarithmic and statistical processes for satisfactory illumination balancing. The development aligns with ongoing efforts to create better methods for the digitalization of cultural heritage. The novelty of Log-Stat lies in its two-stage design, which jointly addresses illumination balancing and tonality adjustment in an integrated framework. The first phase redistributes intensity using logarithmic-based approaches infused with statistical measures to brighten under-lit areas and balance the overall illumination. The phase stage adjusts the tonality using tailored statistical methods. The unique combination of mathematical equations in both phases has not been reported in existing algorithms.

When developing Log-Stat, the key objective was to create an expeditious illumination balancing algorithm that can handle different lighting scenarios. Moreover, it is explicitly designed to preserve the structural integrity and tonal consistency of vital archeological features. By leveraging logarithmic and statistical approaches, an integrated framework has been created that mitigates excessive darkness or brightness. This ensures nuanced features and material variations remain visually distinguishable. The design also incorporates color preservation, implemented only on the value channel of the Hue, Saturation, and Value (HSV) color space. This prevented color shifts and preserved saturation fidelity that can misrepresent the intended appearance of the scene. As a result, Log-Stat is principally suitable for the archaeological domain, where precise visualization of details is vital. The intended target audience includes researchers and experts who require an unfailing enhancement method that improves the perceptibility and interpretability of images without compromising the visual accuracy of the content. Intensive tests are conducted, and the vital findings are reported. This study incorporates tailored image processing procedures, enabling the acquisition of more information from captured images using practical solutions. Ultimately, the findings of this research will have a lasting impact on archaeology, benefiting both current and future researchers. This paper is organized as follows: Section 2 reviews the related work; Section 3 describes the proposed algorithm; Section 4 provides the results, comparisons, and required analysis; Section 5 delivers the key conclusions.

2. RELATED WORK

In past years, this topic has been of interest to many researchers, who have developed and introduced various concepts. In 2015, a probabilistic model (PM) was proposed [11], which utilizes PM as a maximum *a posteriori* (MAP) to approximate the illumination and reflectance of the input image in the linear domain. Then, logarithmic transformations are implemented to determine which transformation provides better performance. Accordingly, the MAP model is converted into an energy-minimization domain, and an alternating direction of multiplier model is implemented. In 2016, a fusion-based method (FbM) was introduced [12], which runs a morphological closing process to estimate the illumination part. Next, two versions of the illumination part are produced, representing contrast-enhanced and illumination-enhanced counterparts using the contrast-limited adaptive histogram equalization (CLAHE) method and a sigmoid function. A customized weight is designed for each version, and a multi-scale fusion model is implemented to create an adjusted-illumination part. Lastly, the output image is created by reimbursing the adjusted-illumination part with the reflectance part. In 2017, an algorithm named LIME was presented [13], in which the illumination part is estimated by finding the maximum RGB value for each pixel. The illumination part is refined by using a structure prior model to create the final illumination part. Finally, the refined illumination is used with the reflectance to output the image.

In 2018, another algorithm, called LECARM [14], was offered, which utilizes the response characteristics of cameras. Initially, a proper response model and its variables are set. After that, the illumination part is estimated *via* an exposure-ratio model for every intensity. Using an approximation ratio map, the selected response model is then used to modify every intensity value to a satisfactory exposure. In 2019, a gradient-based method (GbM) was developed [15], focusing on improving the gradients in the dark areas, as they are more sensitive to human vision. In this algorithm, the gradients of the input image are first extracted to be enhanced using a customized method. Next, optional gradient filtering is applied, followed by an image-integration process that utilizes intensity-range constraints to preserve the gradient components while increasing the intensity to a certain amount. In 2020, a semantic-guided method was delivered [16], focusing on harnessing image semantics. The semantic segmentation approach initially obtains the image areas with designated semantics. Next, the semantics are combined and refined with an illumination map estimated from the illumination part. Next, the dark areas are enhanced by guiding the semantic information.

In 2021, a reflection model (RM) was introduced for illumination balancing [17], *via* the aid of the principal-component analysis (PCA) notion. It first stretches the dynamic range of the color image to correct the contrast, then converts the image into the HSV color model. The V component is processed using the multi-scale theory to get the illumination part. Next, the Fechner model is applied to improve the illumination. After that, a PCA-based fusion process is applied to combine the two images. Then, the contrast is enhanced using the CLAHE method. Finally, a transformation back to the RGB color model is attained to create the output image. In 2022, an atmospheric light-based method was proposed [18], keeping it simple by not requiring training or refinement. This method connects the atmospheric scattering and retinex models to adjust the illumination. Moreover, the medium transmission is computed from the saturation information, and the adaptive saturation of scene radiance is approximated using a non-complex approach. In 2023, a sharpening-smoothing method (ShSm) was developed [19], which processes the *V* channel only after converting from RGB into HSV. After that, a multi-scale decomposition process is applied to estimate the details of the sub-images. Next, the CLAHE method is applied to the last estimation for contrast enhancement. Moreover, the details of the sub-images are improved and added to the contrast-enhanced ones. Finally, an RGB image is created by applying the inverse HSV transform, representing the final image.

In the same year, another concept was presented for illumination balancing; namely, the triangle similarity model (TSM) [20]. It works in the HSI domain, specifically the *S* and *I* channels. It implements scaling and translation processes to improve saturation and intensity, while preserving the H component from modifications. It also implements five model-based enhancement procedures to produce images with better-balanced illumination. In 2024, a generative adversarial network was introduced [21], comprising two main networks: the generative network and the adversarial network. The first includes dilated and regular convolutions with max and average pooling, acting as a multi-scale feature extractor to get better feature information. In addition to these two networks, an illumination attention model is employed to reduce feature redundancy by assigning higher weights to significant features. An upgraded loss function is added to decrease color distortions in this context. Many of the reviewed methods utilize CLAHE as an enhancement module. It works by splitting a given image into small, non-overlapping tiles, then applying histogram equalization to each tile to reallocate its pixel intensities. Next, each tile's histogram is clipped using a predefined clip limit, and the surplus pixels are reallocated evenly across all histogram bins to prevent intensity over-amplification and suppress noise. Finally, a soft-interpolation process is applied between neighboring tiles to avoid block artifacts [22].

In recent years, attention has turned to the development of learning [23], coarse-to-fine [24], or AI-based [25] algorithms. One of the branded learning-based frameworks in this area is the LightenNet, which was presented in 2018 [26]. It is a convolutional neural network (CNN)-based method that adopts a retinex-inspired design. In this method, the network learns to approximate the illumination rather than performing direct processing. This approach enables better illumination balancing with minimized artifact generation. Another neural-network framework named CIE-XYZ Net was introduced in 2022 [27]. It is designed to map images back to a more meaningful representation, operating under consistent assumptions of color and illumination. This approach bridges the gap between publicly available processed images and raw ones. In 2023, a method named Retinexformer was introduced [28]. It is a deep learning-based algorithm that utilizes a one-step Retinex framework (OSRF) instead of multi-step frameworks. OSRF initially approximates the illumination components to enhance the dimmed regions and restore the corrupted parts, delivering the output image. This includes the utilization of an illumination-guided transformer (IGT) that employs illumination information to guide the modeling of non-local connections found in non-uniformly-lit regions. Simply put, Retinexformer is obtained by plugging IGT into OSRF.

In 2024, ConvIR was introduced [29], a lightweight CNN-based algorithm for image restoration. It relies on the CNN architecture to learn end-to-end mapping from degraded sets to clean targets using a convolution feature-extraction scheme. Despite its simplicity, it achieved competitive performances. Still, its effectiveness is based on the alignment between real-world scenarios and training data. In 2025, UPT-Flow, a multi-scale transformer, was presented [30]. This method utilizes a learning transformer backbone to model intricate allocations of intensities, allowing controlled mapping to generate illumination-corrected results. Unlike conventional learning-based approaches, this method uses a probabilistic-learning framework that preserves structural information when adjusting for illumination.

As observed from the reviewed methods, various concepts have been developed, ranging from simple to complex, standard to AI-based. Despite significant advances in this field, not all the introduced algorithms are perfect, as some introduce artifacts, such as halos and distortions, others are of high complexity, and others are inapplicable for real-life applications. Thus, existing algorithms still struggle under non-ideal lighting conditions. Classical algorithms, such as Retinex or histogram equalization, cannot be used, as they provide blind global processing without considering the spatial context and delicate details.

Their improved counterparts mitigated these issues, but are still struggling in this context, as artifacts may still be introduced along with deficient processing abilities in different scenarios. Moreover, despite the increased use of deep-learning and AI-based methods in recent years, image-processing algorithms remain critical for archeological images due to their computational efficiency, reproducibility, and interpretability. In this context, datasets are often limited, making deep-learning methods less practical, as such methods require extensive training. Besides, they act as a black-box, limiting their reproducibility. Moreover, image-processing methods are preferred in domains such as archeology, because they are mathematically transparent, methodical, and reproducible, ensuring long-term reliability and interpretability. Thus, the door remains open to develop an algorithm that considers the advantages and avoids the disadvantages of the reviewed methods, tailored to the nature of the archaeological images. The developed Log-Stat addressed these issues by utilizing a dual-phase approach for illumination balancing and tonality adjustment, operating in a fully transparent and parameterizable manner. This allowed experts to comprehend and control the processing procedure, thereby enabling interpretability, reproducibility, and tracability. Compared to classical, improved, and deep-learning methods, Log-Stat delivers a principled, transparent, fully explainable solution customized to the subtle challenges of archaeological lighting conditions.

3. PROPOSED ALGORITHM

Balancing uneven illumination after image capturing is uneasy and requires customized algorithms to achieve this task successfully. Hence, a tailored algorithm named Log-Stat is developed to balance the inconstant illumination. The name “Log-Stat” signifies two main aspects: logarithmic operations (Log) and statistical methods (Stat). This means that the proposed algorithm relies heavily on these aspects to balance the illumination of a given image. The Log-Stat algorithm starts by changing the input image to the HSV model and processing only the value channel V (being in a 0 to 1 range) while not modifying the other channels. All the equations receive images with a zero-to-one range, avoiding issues, such as under- or overflow during calculations. This aligns well with floating-point arithmetic, which is a prevalent practice in modern processing frameworks. Although the HSV is a non-perceptual color model, it is preferred over color models, such as CIELAB, because it provides a more direct way to enhance the illumination without affecting the colors. In HSV, the V channel corresponds directly to intensity, while in CIELAB, for example, improving the luminance channel L can alter the chromatic components and unintentionally affect the colors. Moreover, HSV is computationally more effective, making it a desirable choice for many real-life applications. The algorithm receives channel V and parameter γ, a numerical value responsible for the illumination level, such that $\gamma > 1$, and a higher value leads to less illumination and further adjusted tonality. Channel V is first processed using Eq. (1) [31]:

$$I_1 = \log\left(V + \varepsilon 1\right) \tag{1}$$

where I_1 is the resulting channel, and $\varepsilon 1 = 0.001$ is a small value to avoid the log of zero, which is infinity. The log transform expands the intensity range to improve visual details in dark regions while compressing the intensity range in brighter areas. This makes it particularly useful for balancing uneven illumination phenomena. Accordingly, a single log transform may help with illumination balancing. Still, it is frequently insufficient individually, especially for archaeological images where lighting situations vary substantially, as it applies a unified effect for all intensity values, which is not ideal for uneven lighting. Thus, a different log transformation is used for bright and dark areas to achieve a more balanced enhancement. Different transformers were found when searching for another log transformation that can contribute to providing better adjustments. One transform of interest is the transform given in [32], which is simple and better compresses higher values while expanding lower values. The second transformer is expressed in the following manner:

$$I = \frac{\max(V)}{\log(\max(V)+1)} \cdot \log(V+1) \tag{2}$$

where · is a multiplication operator, and *max* denotes the highest value in the array. The log transformation in Eq. (2) must be adapted to the nature of the research problem to perform a better job of illumination balancing. Hence, this transformer was heuristically modified to become:

$$I_2 = \frac{-\sigma(V)}{\log(\sigma(V)+1)} \cdot \log((0.925-V)+\varepsilon 2) \tag{3}$$

where σ represents the standard deviation of V and $\varepsilon 2 = 0.09$. This transformer achieves the following points. First, it provides better compression for high values and boosts low values, making details in shadows more noticeable. Second, it reverses the intensities, changing white to black and black to white, acting as a map, where the dark spots exist and are converted into white for addition in the next step, thereby achieving better illumination balancing. The standard deviation is used, as it adapts based on the image tonality, meaning that the effect changes depending on the statistical properties of each input image. Figure 1 provides an intuitive visual demonstration of why to use Eq. (3) instead of Eq. (2) in Log-Stat.

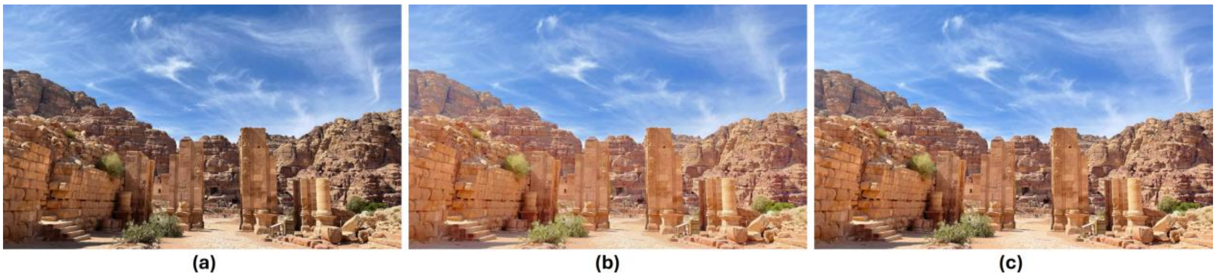


Figure 1. Intuitive visual demonstration of utilizing Eq. (3): (a) Under-lit image; (b) Using Eq. (3) with Log-Stat; (c) Using Eq. (2) with Log-Stat.

Next, the outputs of Eq. (1) and Eq. (3) are combined using a logarithmic image-processing (LIP) addition model. The LIP addition models combine two images to provide a more meaningful output image. Different models exist in this context, and one model of interest is the one introduced by Jourlin and Pinoli [33], which can be expressed as:

$$L = (I_1 + I_2) + (I_1 \cdot I_2) \tag{4}$$

As simple as it may seem, it has drawbacks. The term ($I_1 \cdot I_2$) causes a noticeable increase in intensity values, especially the high-intensity ones, leading to over-brightening the bright areas. Thus, this term is replaced, and the equation is modified to become:

$$I_3 = (I_1 + I_2) \cdot (1 - I_1)^{0.1} \tag{5}$$

In this context, this model performs a non-linear addition process, enhancing darker areas more effectively. The term $(1\text{-}I_1)^{0.1}$ compresses high-intensity values, ensuring that brightness does not unnecessarily increase, which occurs in standard LIP models. The LIP model in Eq. (5) improves the illumination in the dark regions without extremely brightening the bright areas. Figure 2 demonstrates the sensitivity analysis for Eq. (5).

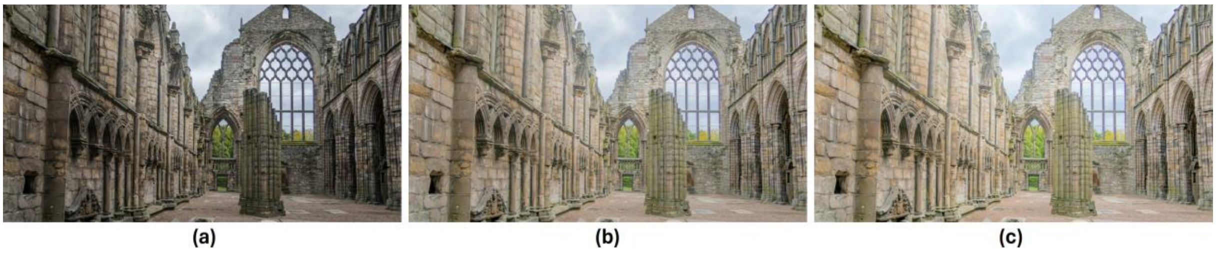


Figure 2. Sensitivity analysis for Eq. (5): (a) Under-lit image; (b) Exponent = 0.1; (c) Exponent = 1.

As observed in Figure 2, the use of exponent = 0.1 in Eq. (5) helped deliver natural overall illumination z and did not increase brightness, especially in the cloud area in the sky. However, when employing exponent = 1, the overall illumination showed a slight increase. Moreover, the brightness in the cloud area was also increased, which is undesirable. Next, three more models are utilized to balance illumination, which are the following ones:

$$I_4 = 1 - \left(\frac{I_3 - \min(I_3)}{\max(I_3) - \min(I_3)} \right) \tag{6}$$

$$I_5 = 1 - (I_4 - I_1) \tag{7}$$

$$I_6 = \frac{I_5 - \min(I_5)}{\max(I_5) - \min(I_5)} \tag{8}$$

Eq. (6) performs statistical normalization [34] and inversion. Normalization redistributes I_3 values to the full range of zero to one, ensuring that the image intensities fall within the proper image range. Subtraction from one makes dark areas bright, and *vice versa*, reversing the action of inversion performed in Eq. (3). As for Eq. (7), it alters I_4 by subtracting I_1 to ensure the primary illumination structure is maintained, harmonizing the illumination-balancing process, so that the output does not drift too far from the input, and ensuring that the balancing effect is not over-darkened or over-brightened to provide better naturality. In addition, this step is not purely subtractive, but it reconstructs a compensated-illumination image by reintegrating structural information derived from I_3. Eq. (8) rescales the intensities of I_5 to a valid image range [34]. At this point, the illumination is balanced, illuminating the dark regions while maintaining bright regions from being over-illuminated. Still, the totality of the output image I_6 is not well-adjusted, and it appears to have foggy and washed-out effects. Thus, a tonality-enhancement model is applied to I_6, which includes four distinct steps, expressed as:

$$T_1 = 1 - \exp\left(-\frac{I_6}{\sqrt{I_6}} \right) \tag{9}$$

$$T_2 = \frac{1}{1 + \left(\frac{I_6}{\gamma} \right)^{-\gamma}} \tag{10}$$

$$T_3 = \left(T_1 + (T_1 \cdot T_2) \right)^{\gamma} \tag{11}$$

$$T_4 = \frac{T_3 - \min(T_3)}{\max(T_3) - \min(T_3)} \tag{12}$$

Eq. (9) is a modified Rayleigh cumulative distribution function (MRCDF) [35], a statistical method that acts as a curvy transformation and can enhance the tonality by non-linearly mapping the intensities. It modifies the dynamic range by increasing lower values, but relatively unchanging higher ones. Eq. (10) is the gamma-adjusted CDF of the log-logistic distribution (GA-CDF-LLD) [36], a modified statistical method used to improve mid-tones while preserving very bright and very dark regions, with the help of the γ parameter, in that higher values boost mid-tones. Eq. (11) is the modified LIP model given in Eq. (4), which combines the output of Eq. (9) and that of Eq. (10), since a single mathematical method cannot achieve the task of tonality enhancement. This equation performs an overall-tonality refinement based on the value of the γ parameter, providing an adaptive increase in the difference between the lowest and the highest intensities, attenuating the foggy effect. The T_2 term is missing from the additive part in Eq. (11), because when adding it, it slightly dims the overall brightness. That's why it is removed to allow better overall-illumination representation and to reduce computations. Eq. (12) is the final equation, which is the statistical normalization [34] applied to ensure that no intensities fall outside the display range, preventing loss of dynamic range and improving the overall tonality of the output. The output of Eq. (12) is the enhanced value channel. Thus, a conversion into the RGB color model is applied, generating the output image.

Regarding the coupling role in Eqs. (10) and (11), an ablation study has been conducted to prove the effectiveness of coupling. Accordingly, γ was used only for Eq. (10), while fixing it for Eq. (11). The results in Figure 3 reveal that decoupling led to inconsistent and unstable behavior with different γ

values. Hence, low values (γ=1-3) under-enhanced the image and were unsuccessful in recovering acceptable visibility. Values over 7 (γ=10-100), on the other hand, did not affect the output; it's like that the algorithm lost its adjustment control. This instability arose, because these two equations jointly regulate the tonal adjustment. When they are decoupled, the two equations respond in conflicting ways, which makes it unpossible to uphold a balanced enhancement. In contrast, the coupling strategy allowed intrinsic synchronization between fusion strength and tonal adjustment. As a result, natural, stable, and visually coherent results are obtained when the coupling strategy is utilized. These findings rationalize the design choice of enabling the coupling role of γ to coordinate the two operations and verify that the coupling process is vital rather than arbitrary.

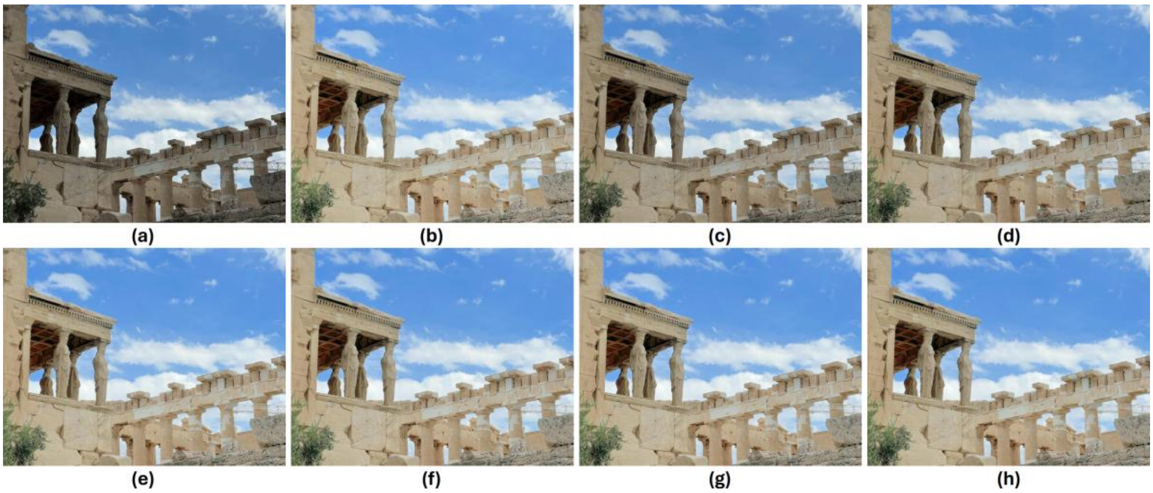


Figure 3. Ablation study (justifying the coupling effect): (a) Under-lit image; (b) Standard setting: coupling effect with γ = 7; Decoupling effect using fixing exponent for Eq. (11) at 7 and γ is equal to: (c) γ = 1; (d) γ = 3; (e) γ = 10; (f) γ = 20; (g) γ = 50; (h) γ = 100.

The diagram of the Log-Stat algorithm is given in Figure 4. In sensitive areas, such as archaeology, classical image-processing frameworks are often chosen, because their operations are transparent and methodically reproducible. They apply mathematically-defined approaches, permitting one to understand exactly how each step modifies the image. This clarity helps uphold subtle features without introducing artifacts or unrealistic details, which can occur in AI or deep learning-based methods. Classical frameworks also allow accurate parameter control, making it easier to adapt processing to uneven-illumination challenges, while ensuring that the outcomes remain true to the original scene. Thus, classical transformations have been utilized with Log-Stat to achieve these targets.

4. RESULTS AND ANALYSIS

This section explains different aspects related to the outcomes. The dataset was collected from https://unsplash.com/, one of the largest photo-stock websites online, containing millions of uncopyrighted images that can be used free of charge. A thorough search was conducted on this website, and 200 images were collected, representing a diverse range of archaeological scenes. In this paper, no dedicated pre-processing was carried out on the dataset images. This is deliberate, because the purpose is to evaluate the general applicability and reliability of Log-Stat when used with real-world archeological images without any form of enhancement. As for the image-selection procedure, it was conducted in a semi-random manner from the collected images. Moreover, this procedure is implemented without imposing restrictions on the image content or the severity of lighting. This tactic ensures that the assessment truly reflects natural and diverse occurring image-lighting conditions rather than a curated selection tailored to Log-Stat strengths. It is also worth noting that no unified dataset has been used, as the intention was to evaluate performance under challenging illumination conditions in real-world scenarios. Such conditions vary widely and cannot be fully signified by a standardized dataset. Thus, diverse real-world examples were employed intentionally to capture the variability and complexity of archaeological scenes. This allows a more meaningful assessment of Log-Stat’s practical effectiveness. Also, it shows its ability to handle challenging illumination conditions posed by the scene environments.

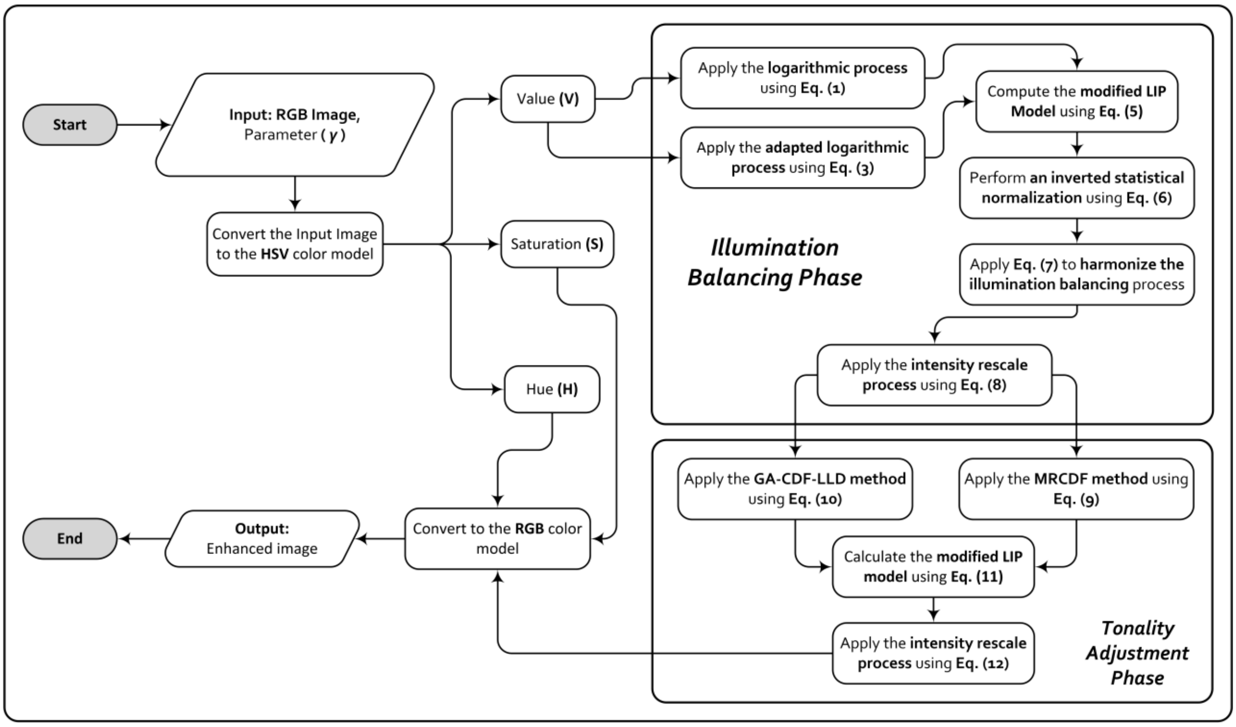

Start
Input: RGB Image, Parameter (γ)
Convert the Input Image to the HSV color model
Value (V)
Saturation (S)
Hue (H)
Apply the logarithmic process using Eq. (1)
Apply the adapted logarithmic process using Eq. (3)
Compute the modified LIP Model using Eq. (5)
Perform an inverted statistical normalization using Eq. (6)
Apply Eq. (7) to harmonize the illumination balancing process
Illumination Balancing Phase
Apply the intensity rescale process using Eq. (8)
Apply the GA-CDF-LLD method using Eq. (10)
Apply the MRCDF method using Eq. (9)
Calculate the modified LIP model using Eq. (11)
Apply the intensity rescale process using Eq. (12)
Tonality Adjustment Phase
Convert to the RGB color model
Output: Enhanced image
End

Figure 4. Illustrative diagram of the Log-Stat algorithm.

The experimental results are divided into three categories. The first category describes the effects of changing the value of γ. It is provided to offer a visual understanding of what γ does to the image when it changes, with sample results shown in Figure 5. The second category demonstrates the ability to process noisy images, and the outcomes are shown in Figure 6. The third category is processing various archaeology-related images and showing the before and after versions to demonstrate Log-Stat’s capabilities. The results of this action are shown in Figures from 7 to 9. Besides, the proposed algorithm is compared with ten contemporary algorithms. The comparison methods are PM, FbM, LIME, LightenNet, LECARM, GbM, RM, ShSm, TSM, and Retinexformer, which are reviewed in the related work section. All experiments and comparisons are performed using a laptop equipped with an i5-1135G7 2.40 GHz processor and 16 GB of RAM. The comparison results are given in Figures from 10 to 13 and Tables from 1 to 6. The image sizes of Figures from 10 to 13 are: (6240×4160), (3200×1975), (4032×3024), and (4274×3205), respectively. In this context, an insightful analysis of the results of each category is given. Then, this section ends with key remarks. For performance measurement, six measures were used, which are: lightness order error (LOE) [37], blind multiple pseudo-reference index (BMPRI) [38], color-quality measure (CQM) [39], perception-based image-quality evaluator (PIQE) [40], gradient magnitude with Laplacian of Gaussian(GM-LOG) [41], and runtime (RT) [42].

LOE is a reduced-reference metric used for quantifying the relativity of lightness-order, meaning how much the brightness grade among neighboring pixels is preserved after enhancement. This reflects illumination's naturalness and perceptual consistency, in that if LOE is significantly changed, the enhanced image appears artificial or unnatural. BMPRI is a no-reference metric that measures naturalness and artifact presence, aiming to reflect how humans perceive the image. CQM is a no-reference metric that measures color quality using three perceptual attributes: contrast, colorfulness, and sharpness, aiming to mimic human perception of observing colors. PIQE is a no-reference metric that evaluates how a given image diverges from undistorted and natural visual features. It measures perceptual distortions. GM-LOG is also a no-reference metric that evaluates a given image based on the statistics of gradient information. It assumes that high-quality natural-looking images follow foreseeable gradient patterns, because they are highly structured, and degradations distort these patterns and deform the structure. It measures structural naturalness. RT is a measure used to quantify the computational cost of given algorithms, which is essential to assess their practical usability. Even if the algorithms are based on dissimilar concepts, RT delivers insight into computational efficiency and suitability for real-life or resource-limited applications. For LOE, PIQE, and RT, lower scores indicate better performances, meaning that the enhancement preserved more natural illumination relations for LOE, fewer perceptual distortions for PIQE, and faster implementation for RT. In contrast, higher scores indicate better performances for BMPRI, CQM, and GM-LOG, suggesting more perceptual details, higher color quality, and better structural naturalness.

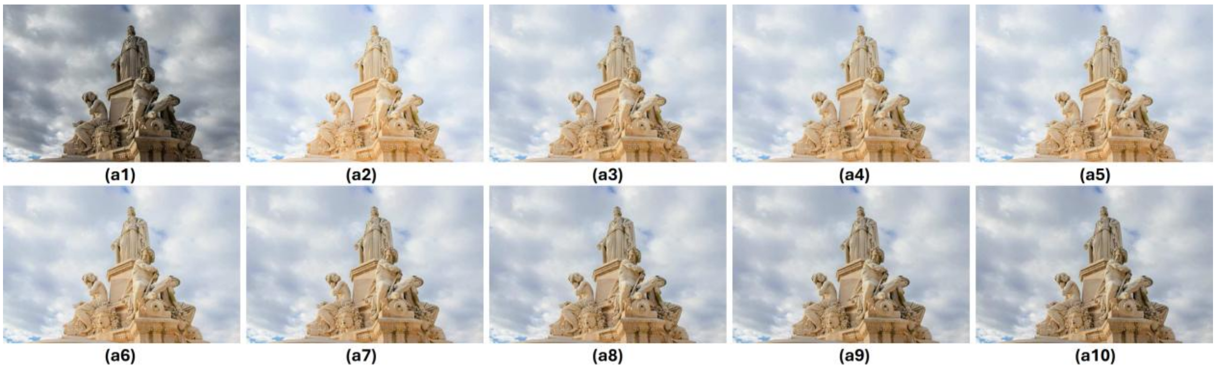


Figure 5. The impact of changing γ. (a1) unevenly-illuminated image. Log-Stat processed other images with γ equal to: (a2) 1; (a3) 2; (a4) 3; (a5) 4; (a6) 5; (a7) 6; (a8) 7; (a9) 8; (a10) 9.

The results in Figure 5 illustrate the effect of changing γ, where the brightness decreases when γ increases in value, and the tonality further improves. This indicates that there is an inverse relation between brightness and γ, while the relationship is direct when it comes to tonality. Moreover, it is observed that the visibility of the details, specifically in the highlights and shadows, is enhanced with mediocre γ values. The image appears overly illuminated around (a2) to (a3), which may lead to the loss of certain fine details and an unnatural appearance. In addition, as γ increases, the image progressively darkens, balancing the illumination and enhancing the tonality. It is noticed that a moderate γ value (around 4-7) appears to provide a well-balanced illumination without over-exposure. Thus, the choice of γ is image-dependent based on specific features and the required illumination boost level. In this field, it is often preferred to use a manual enhancement parameter rather than an automatic one, as such images present unique and highly variable challenges that automation may not address effectively. Shadows, uneven lighting, and irregularities can confuse automated algorithms, leading to over- or under-correction. In contrast, manual adjustment enabled fine-tuning, allowing higher accuracy and improved extraction of meaningful information. Figure 6 shows the performance of Log-Stat when applied to noisy images. The original images shown in (a1, b1) exhibit noticeable illumination non-uniformity, with their zoomed-in views (a2, b2) revealing pronounced noise and texture degradation. After processing with Log-Stat at two γ settings (6 and 10), the resulting outputs (a3, b3, a4, b4) show notable improvement in tonality and illumination balance.

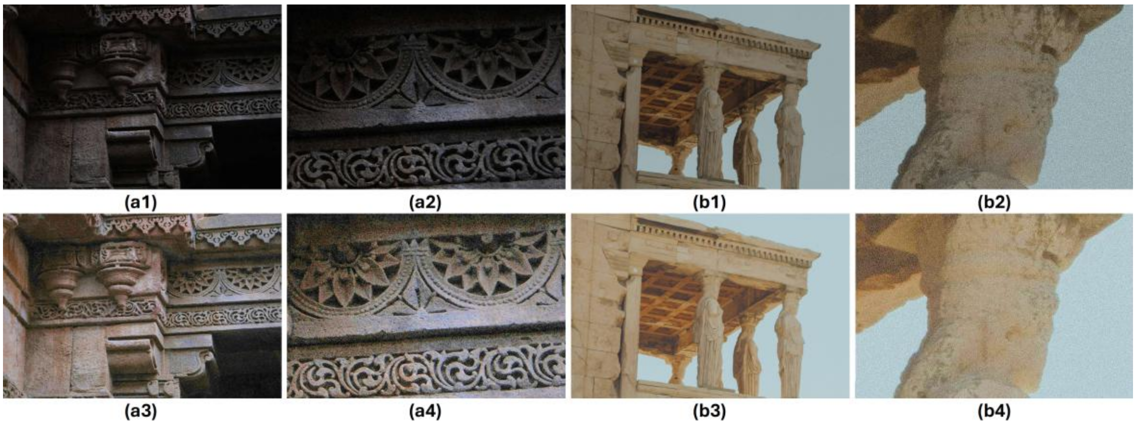


Figure 6. Log-Stat results with noisy images. (a1, b1) unevenly-illuminated images. (a2, b2) zoomed-in regions of (a1, b1) images. (a3, b3, a4, b4) processed by Log-Stat with γ = 6 and 10.

While some noise became more visible, especially in darker and shadowed regions, the structural details and textures have been improved. This indicates that the algorithm upholds its enhancement capability under noise conditions, though higher γ values may reveal more noise. Despite this, the algorithm does not introduce edge artifacts or artifacts, keeping the enhancement process structurally reliable even in noisy conditions. Moreover, despite mild noise revelation with higher γ in some images, this behavior is anticipated and does not undermine the purpose of Log-Stat. This is due to it being designed exclusively for illumination balancing, not for denoising. Thus, it processes under the principle of maintaining the pristine image features while balancing illumination. Accordingly, when the brightness is increased in the dark regions, the formerly hidden noise logically becomes more visible. This does not mean that Log-Stat introduces or increases noise, but means that it uncovers existing noise that was concealed by low brightness. This result is consistent with most illumination-balancing algorithms that do not utilize a specialized denoising model. More importantly, the goal of Log-Stat is to achieve consistent, natural, and accurate illumination balancing without amplifying existing noise or introducing artifacts, which is attained successfully. Integrating a denoising model would risk altering image details, which is not desirable for archaeological images.

From the experimental results in Figures from 7 to 9, an overall improvement in quality is observed, as the processed images have balanced illumination and improved tonality compared to the original versions. Moreover, the shadows in certain areas are attenuated, making details in the dark regions more visible. Likewise, the fine surface textures appear more distinct and clearer. For example, in the images containing carvings in Figures 9e1 and 9e2, the sophisticated features of the sculptures are more pronounced after filtering. As for colors, the filtered images preserve natural color tones while handling the undue bright or dark regions, and the Log-Stat does not introduce color distortions.

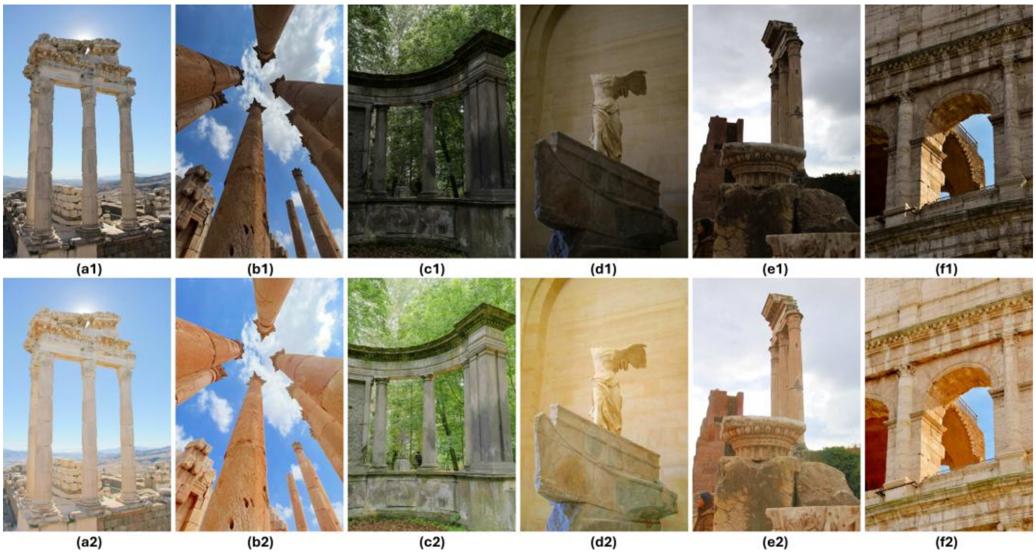


Figure 7. Log-Stat results (Set 1). (a1-f1) Images with imbalanced illumination. (a2-f2) Processed with γ = (3.5, 4, 6, 5, 5, and 4.5).

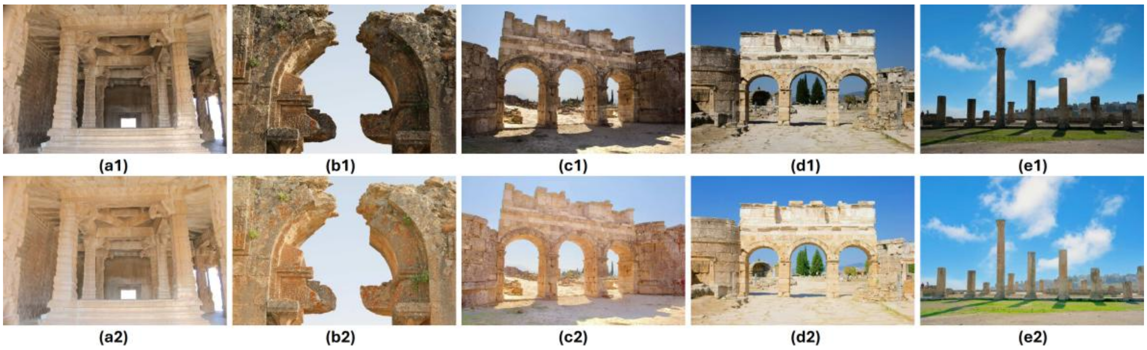


Figure 8. Log-Stat results (Set 2). (a1-e1) Images with imbalanced illumination. (a2-e2) Processed with γ = (3, 3.5, 6, 6, and 6.5).

As for shadows, original images contain sturdy shadows due to inconsistent lighting conditions. Still, Log-Stat has weakened such effects, delivering a pleasing appearance across the structures. This indicates that it effectively corrects these under-exposed areas, making formerly dark areas more evident. For example, the inner parts of the temple in Figures 8a1 and 8a2, which were dark, are now brighter and display more visible structural details. In Figures 8c1 and 8c2, the rock archway had a darkened foreground. Now, the structure appears uniformly bright. As for illumination, it seems natural and avoids unnecessary over-exposure, ensuring that details are retained while upholding balance between tonality and texture. Likewise, a better color balance is observed, especially with warm tones (sandy and brown). Other colors, such as the blue in the sky, appeared more vibrant without looking unnatural. Moreover, it did not introduce major color distortions. Such enhancements have improved the overall scene clarity, achieving the key objective of the algorithm.

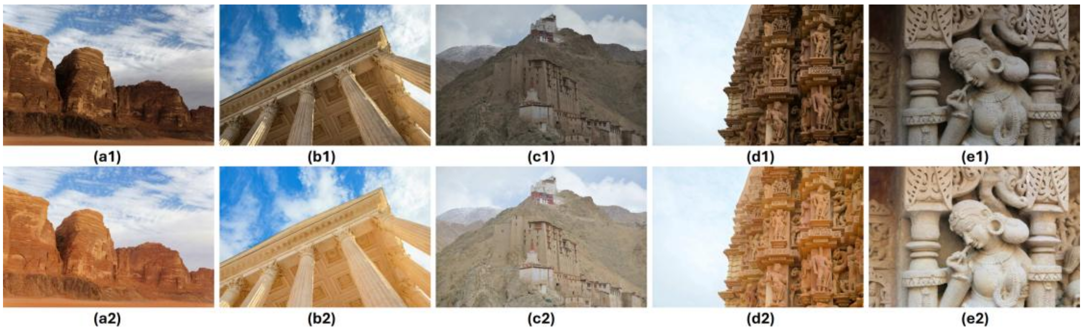


Figure 9. Log-Stat results (Set 3). (a1-e1) Images with imbalanced illumination. (a2-e2) Processed with γ = (6, 5, 1.5, 6, and 5.5).

Although certain processed images may appear visually similar to over-exposure, this is an unintended brightness increase, but in fact, it is the effect of illumination balancing. The intended illumination-balancing process should brighten shadowed areas and adjust the illumination across the scene. This makes previously obscured surfaces appear brighter and more uniformly illuminated. The illumination balancing may give certain regions a washed-out appearance when compared to their dimmed counterparts, but in fact, this is the balancing result that makes certain surfaces appear brighter, even though intensities remain within a valid range with no saturation occurrence. This also reflects Log-Stats's emphasis on revealing hidden details rather than preserving pristine non-uniform lighting. To be more specific, the textural information is not lost or degraded. However, it becomes more uniformly lit once shadows are attenuated and under-illuminated areas are brightened.

The comparison results are shown in Figures from 10 to 13. Tables from 1 to 6 demonstrate the dissimilarity of each algorithm in illumination balancing. All have improved the illumination, each in its way. The measures used indicate the following: LOE (illumination naturalness), RT (execution speed), BMPRI (perceptual clarity), CQM (color quality), PIQE (perceptual distortions), and GM-LOG (structural naturalness).

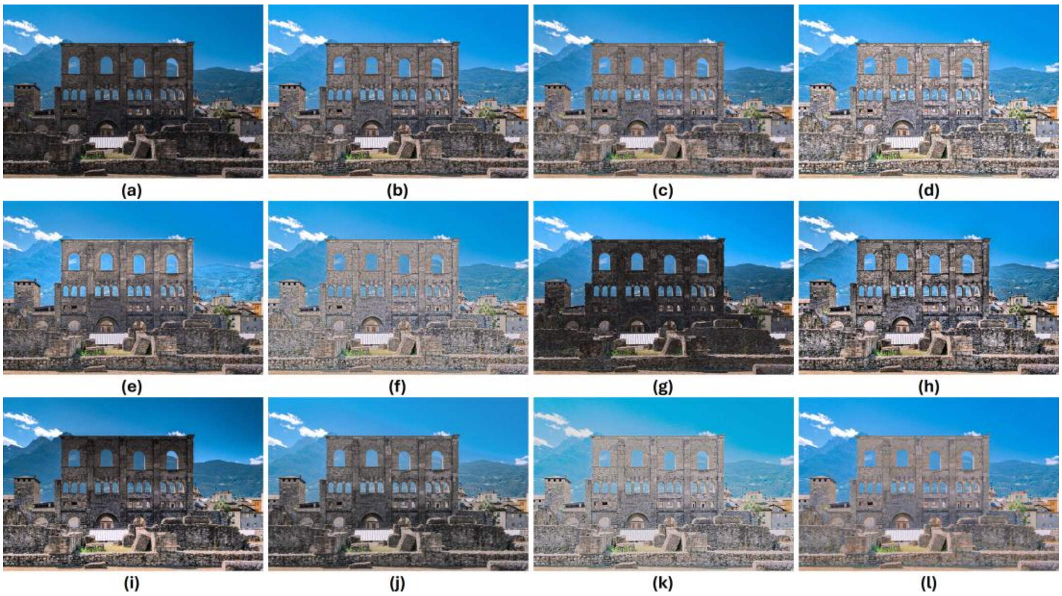


Figure 10. Results of comparisons (Batch 1): (a) Original image; images (b-l) are processed by: (b) PM; (c) FbM; (d) LIME; (e) LightenNet; (f) LECARM; (g) GbM; (h) RM; (i) ShSm; (j) TSM; (k) Retinexformer; (l) Log-Stat.

The performance ranking is set from worst to best as follows: lowest, below-low, low, above-low, below-mediocre, mediocre, above-mediocre, below-high, high, above-high, and best. The analysis is based on these attributes, ranks, and the detected drawbacks. Accordingly, PM provided sub-optimal illumination with slight brightness amplification in the bright areas. Thus, it scored above high in LOE, BMPRI, CQM, and GM-LOG, but below mediocre in PIQE, while being the 7th fastest in RT, recording an average of 48.9 seconds. The FbM introduced shadows to certain image regions due to the use of the standard CLAHE method. This justifies scoring below high in LOE, lowest in BMPRI, and mediocre in CQM, above high in PIEQ, and low in GM-LOG, while ranking as the 6th fastest in RT, with an average time of 9.8 seconds.

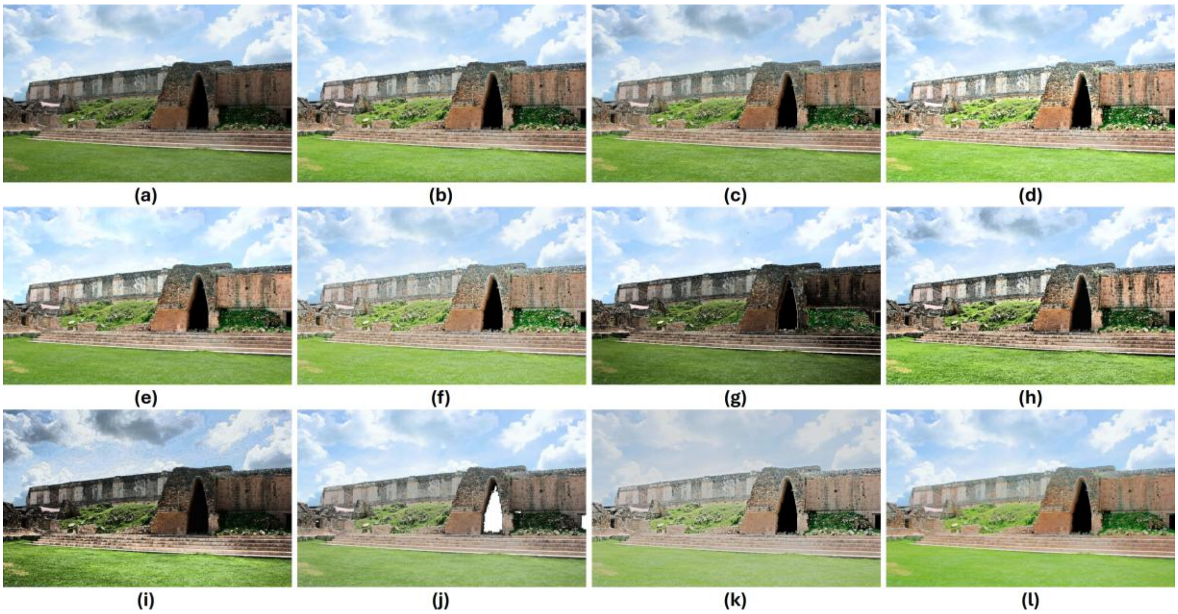


Figure 11. Results of comparisons (Batch 2): (a) Original image; images (b-l) are processed by: (b) PM; (c) FbM; (d) LIME; (e) LightenNet; (f) LECARM; (g) GbM; (h) RM; (i) ShSm; (j) TSM; (k) Retinexformer; (l) Log-Stat.

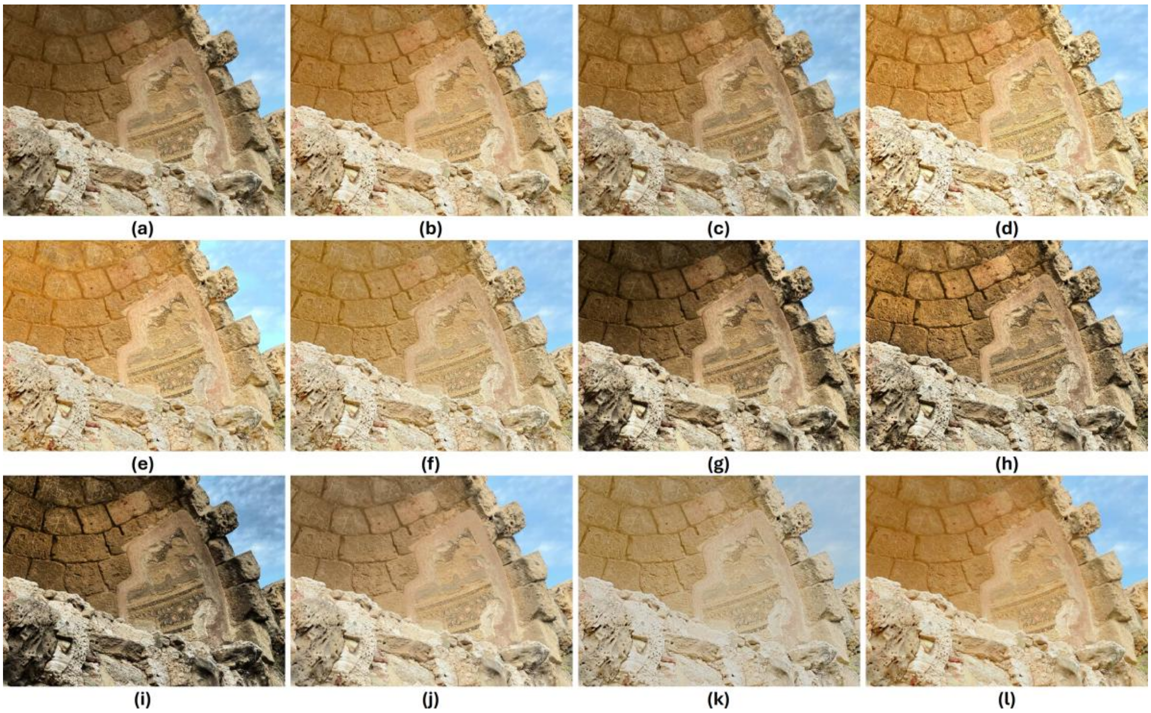


Figure 12. Results of comparisons (Batch 3): (a) Original image; images (b-l) are processed by: (b) PM; (c) FbM; (d) LIME; (e) LightenNet; (f) LECARM; (g) GbM; (h) RM; (i) ShSm; (j) TSM; (k) Retinexformer; (l) Log-Stat.

LIME, on the other hand, provided unnatural highlights and brightness amplification in certain regions with brightened colors. This is why it attained the lowest in LOE, above mediocre in BMPRI, below high in CQM, low in PIQE, and below mediocre in GM-LOG, with a relatively fast implementation, ranking the 2nd fastest in RT, with an average speed of 5.2 seconds. The LightenNet, on the other hand, introduced a haloing effect around edges and unnatural illumination, especially in dark scenes. This justifies its scoring, low in LOE, below mediocre in BMPRI, above mediocre in CQM, below high in PIQE, and high in GM-LOG. Speaking of RT, it was the slowest comparison method, operating at an average of 188.1 seconds. In contrast, LECARM delivered abnormal illumination in the dark areas, scoring below mediocre in LOE and below low in BMPRI, high in CQM and PIQE, and mediocre in GM-LOG, while averaging at 8.1 seconds in RT, placing 5th rank.

As for GbM, it introduced illumination errors, insufficient illumination, and distortions. This explains the recorded scores above mediocre in LOE and GM-LOG, low in BMPRI, while scoring the lowest in CQM and PIQE, ranking as the 9th slowest algorithm in the comparison. RM produced shadows and distortions due to the use of the standard CLAHE method, which rationalizes the performances of above low in LOE and CQM, mediocre in BMPRI and PIQE, and lowest in GM-LOG, while placing 4th in RT. ShSm also introduced shadows and distortions because of CLAHE utilization, but with varying levels. This justifies ranking mediocre in LOE, high in BMPRI, below low in PIQE, and above low in PIQE and GM-LOG. In terms of RT, it was the 3rd fastest. The TSM generated distortions (Figure 11i) and insufficient illumination. Yet, it scored high in LOE, above low in BMPRI, below mediocre in CQM, above mediocre in PIQE, and below high in GM-LOG, placing 8th rank in RT, averaging 50.8 seconds. Retinexformer provided varying performances as the contrast was deficient, the colors were slightly pale, and the whiteness in the resulting images tended to be yellowish. This explains scoring below low in LOE, below high in BMPRI, low in CQM, and below low in PIQE and GM-LOG. As for RT, this method was the second slowest among the competitors. The proposed Log-Stat algorithm performed the best in all measures. This is a significant advancement as the resulting images by Log-Stat have natural illumination, high perceptual clarity, vivid colors, minimal visual distortions, and are obtained in a few seconds. These qualities are important for accurate archaeological findings, especially when images are captured in uncontrolled or non-ideal lighting conditions. Moreover, Log-Stat achieves this efficiently, processing large images in an average of just 2.4 seconds. This rapid execution makes the method highly practical for real-time or on-site applications. The combination of visual clarity and computational speed underscores the Log-Stat's potential for integration into post-processing workflows and fieldwork.

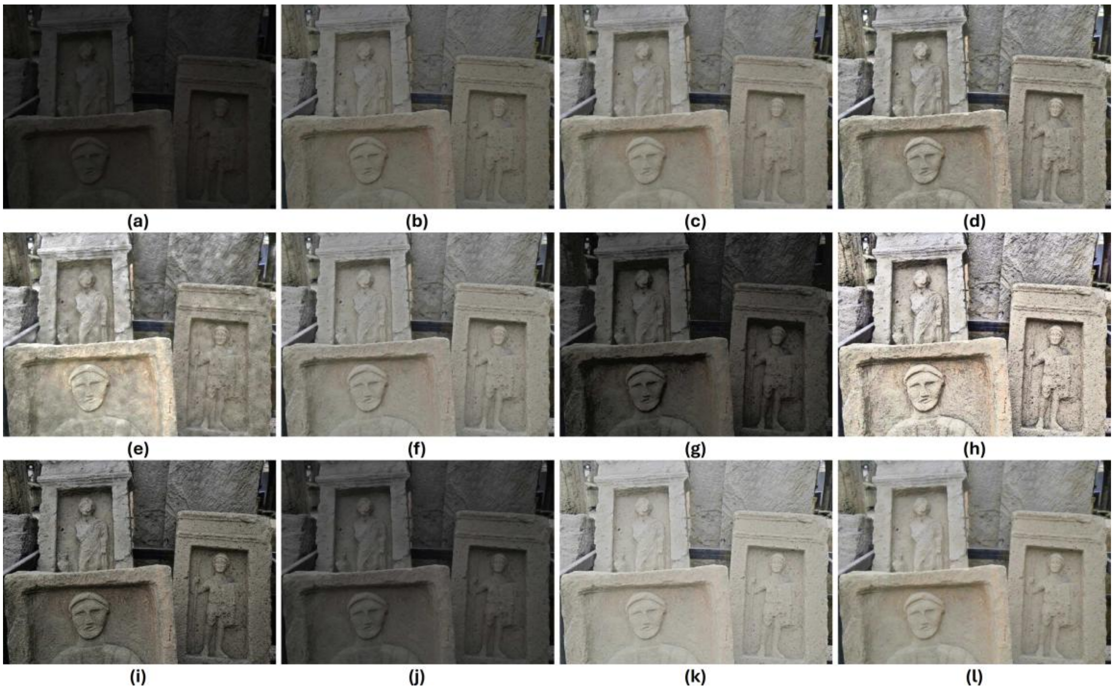


Figure 13. Results of comparisons (Batch 4): (a) Original image; images (b-l) are processed by: (b) PM; (c) FbM; (d) LIME; (e) LightenNet; (f) LECARM; (g) GbM; (h) RM; (i) ShSm; (j) TSM; (k) Retinexformer; (l) Log-Stat.

Table 1. RT scores ↓ of the comparison.

#	Algorithms	Figure 10	Figure 11	Figure 12	Figure 13	Average	Rank
1	PM	102.824	26.165	27.086	39.589	48.916	7th
2	FbM	21.120	3.548	6.900	7.731	9.824	6th
3	LIME	9.312	2.380	4.537	4.716	5.236	2nd
4	LightenNet	382.751	72.583	135.895	161.308	188.134	11th
5	LECARM	14.285	3.330	6.491	8.373	8.119	5th
6	GbM	148.972	39.516	40.288	38.310	66.771	9th
7	RM	15.047	2.535	5.677	5.766	7.256	4th
8	ShSm	10.867	2.152	4.202	4.826	5.511	3rd
9	TSM	136.140	15.749	46.460	5.212	50.890	8th
10	Retinexformer	308.582	75.814	133.286	61.012	144.673	10th
11	Log-Stat	4.349	1.070	2.0487	2.147	2.403	1st

Table 2. LOE scores ↓ of the comparison.

#	Algorithms	Figure 10	Figure 11	Figure 12	Figure 13	Average	Rank
1	PM	112.653	247.252	194.801	149.699	176.101	2nd
2	FbM	439.646	435.024	348.447	264.396	371.878	4th
3	LIME	1058.800	1280.200	1134.100	475.512	987.153	11th
4	LightenNet	672.504	702.240	1020.300	597.401	748.111	9th
5	LECARM	746.188	741.846	690.645	291.088	617.441	7th
6	GbM	483.855	655.782	359.066	382.355	470.264	5th
7	RM	584.585	708.174	645.666	601.127	634.888	8th
8	ShSm	735.189	572.511	407.902	366.432	520.508	6th
9	TSM	154.698	434.263	169.655	119.474	219.522	3rd
10	Retinexformer	531.314	1419.600	1041.400	367.760	840.018	10th
11	Log-Stat	107.039	204.899	159.521	113.901	146.340	1st

Table 3. BMPRI scores ↑ of the comparison.

#	Algorithms	Figure 10	Figure 11	Figure 12	Figure 13	Average	Rank
1	PM	22.816	19.104	13.248	28.201	20.842	2nd
2	FbM	20.713	10.737	11.374	24.824	16.912	11th
3	LIME	22.646	13.114	11.152	29.217	19.032	5th
4	LightenNet	20.594	10.780	11.003	30.091	18.117	7th
5	LECARM	21.576	11.326	11.161	23.966	17.007	10th
6	GbM	20.578	10.621	15.729	24.492	17.855	9th
7	RM	22.130	15.455	13.075	22.724	18.346	6th
8	ShSm	22.312	17.913	13.486	27.271	20.245	3rd
9	TSM	18.355	12.289	10.823	30.896	18.090	8th
10	Retinexformer	19.401	11.888	18.368	29.043	19.675	4th
11	Log-Stat	23.232	21.210	13.585	28.924	21.737	1st

Table 4. CQM scores ↑ of the comparison.

#	Algorithms	Figure 10	Figure 11	Figure 12	Figure 13	Average	Rank
1	PM	0.138	0.203	0.278	0.115	0.183	2nd
2	FbM	0.139	0.188	0.260	0.107	0.173	6th
3	LIME	0.152	0.199	0.286	0.072	0.177	4th
4	LightenNet	0.133	0.208	0.268	0.092	0.175	5th
5	LECARM	0.156	0.187	0.274	0.104	0.180	3rd
6	GbM	0.086	0.103	0.188	0.042	0.104	11th
7	RM	0.140	0.183	0.245	0.103	0.167	8th
8	ShSm	0.098	0.140	0.177	0.062	0.119	10th
9	TSM	0.146	0.210	0.261	0.064	0.170	7th
10	Retinexformer	0.154	0.157	0.213	0.128	0.163	9th
11	Log-Stat	0.165	0.227	0.276	0.123	0.197	1st

Table 5. PIQE scores ↓ of the comparison.

#	Algorithms	Figure 10	Figure 11	Figure 12	Figure 13	Average	Rank
1	PM	24.318	29.515	14.618	23.095	22.886	7th
2	FbM	18.398	30.415	19.509	10.402	19.681	2nd
3	LIME	21.340	31.495	27.001	12.799	23.158	9th
4	LightenNet	19.449	30.554	21.613	13.258	21.218	4th
5	LECARM	18.728	29.462	21.564	11.450	20.301	3rd
6	GbM	20.874	33.646	26.949	13.164	23.658	11th
7	RM	20.461	33.519	22.114	14.898	22.748	6th
8	ShSm	20.162	35.800	22.861	13.726	23.137	8th
9	TSM	19.717	23.841	19.214	22.715	21.371	5th
10	Retinexformer	18.117	28.735	36.079	9.840	23.192	10th
11	Log-Stat	18.261	26.137	11.530	13.280	17.302	1st

Table 6. GM-LOG scores ↑ of the comparison.

#	Algorithms	Figure 10	Figure 11	Figure 12	Figure 13	Average	Rank
1	PM	8.041	9.067	7.581	5.392	7.520	2nd
2	FbM	6.711	7.657	6.200	4.862	6.357	9th
3	LIME	6.788	8.307	6.555	5.584	6.808	7th
4	LightenNet	7.946	8.985	6.818	6.118	7.466	3rd
5	LECARM	6.635	8.545	6.812	5.278	6.817	6th
6	GbM	6.971	8.592	6.671	5.287	6.880	5th
7	RM	6.539	6.663	5.623	5.223	6.012	11th
8	ShSm	7.319	7.968	6.266	5.004	6.639	8th
9	TSM	7.014	8.201	8.542	5.462	7.304	4th
10	Retinexformer	7.183	8.002	5.569	4.576	6.332	10th
11	Log-Stat	7.029	10.115	8.553	5.675	7.843	1st

Despite the disadvantages, noise still appears when the dark parts are enhanced, because it already exists, and balancing the illumination further reveals such hidden noise. Therefore, a fast and efficient denoising method should be utilized to attain better quality results. Illuminance balancing can significantly contribute to archaeological documentation, analysis, and 3D modeling. In this study, unevenly-illuminated archaeological images were utilized to facilitate qualitative comparisons with recognized benchmarks and to ensure objective evaluations with reproducible conditions. The proposed Log-Stat algorithm was developed with these issues in mind, focusing on robust processing with various illumination conditions. Furthermore, because Log-Stat is based on mathematical transformations rather than on data-driven training, it is inherently generalized and not limited to a definite dataset. Improving details in unevenly-illuminated archaeological photographs can enhance the visibility of petroglyphs, inscriptions, eroded carvings, or other related subjects. Moreover, when creating 3D models from images, illumination irregularities can generate artifacts in the model. Thus, balancing illumination before 3D modeling ensures that textures are constant and more realistic.

5. CONCLUSION

This paper introduces a low-complexity illumination-balancing algorithm that integrates logarithmic approaches with statistical methods to balance the inconsistent illumination and adjust the tonality of archeological images. This combination allowed for better revealing of structural details and important textures. The Log-Stat was tested with various real-world scenes, compared with prominent algorithms, and the outcomes were evaluated using six measures. The products of experiments demonstrated substantial enhancements in illumination and visibility, providing more clarity to the resulting images. Likewise, experiments showed Log-Stat's competence to handle various scenes with different illumination levels. The comparisons also showed Log-Stat favorability in various visual aspects and processing speed. This study is beneficial in the field of digitalization, especially in 3D modeling, analysis, or documentation. For example, if 3D models were created with images having inconsistent illuminations, the model would have dark regions, making it less desirable for usage. Follow-up research can aim for automation *via* perceptual features. This can contribute to making Log-Stat fully automated using a tailored approach.

REFERENCES

[1] C. Morgan and H. Wright, "Pencils and Pixels: Drawing and Digital Media in Archaeological Field Recording," Journal of Field Archaeology, vol. 43, no. 2, pp. 136–151, 2018.

[2] O. A. Basheer and Z. Al-Ameen, "Illumination Enhancement of Nighttime Images Using a Regulated Single Scale Retinex Algorithm," Jordanian Journal of Computers and Information Technology (JJCIT), vol. 10, no. 2, pp. 138–151, 2024.

[3] P. Sapirstein and S. Murray, "Establishing Best Practices for Photogrammetric Recording during Archaeological Fieldwork," Journal of Field Archaeology, vol. 42, no. 4, pp. 337–350, 2017.

[4] H. Kaur and N. Sohi, "A Novel Enhancement Method for Colored Rock Art Archaeological Images," Int. J. Adv. Res. Comput. Sci. (IJARCS), vol. 8, no. 7, pp. 1163–1167, 2017.

[5] S. Sylaiou et al., "Redefining Archaeological Research: Digital Tools, Challenges and Integration in Advancing Methods," Applied Sciences, vol. 15, no. 5, p. 2495, 2025.

[6] L. Marchesotti, N. Murray and F. Perronnin, "Discovering Beautiful Attributes for Aesthetic Image Analysis," Int. J. of Computer Vision (IJCV), vol. 113, pp. 246–266, 2015.

[7] M. G. Robinson, Photogrammetry for Archaeological Objects: A Manual, ISBN-10, 1743329830, Sydney, Australia: Sydney Univ. Press, 2024.

[8] S. Kang et al., "Image Intrinsic Components Guided Conditional Diffusion Model for Low-light Image Enhancement," IEEE Trans. Circuits Syst. Video Technol., vol. 34, no. 12, pp. 13244–13256, 2024.

[9] S. Xu, X. Chen, B. Song, C. Huang and J. Zhou, "CNN Injected Transformer for Image Exposure Correction," Neurocomputing, vol. 587, p. 127688, 2024.

[10] N. Singhal, A. Kadam, P. Kumar, H. Singh and A. Thakur, "Study of Recent Image Restoration Techniques: A Comprehensive Survey," Jordanian Journal of Computers and Information Technology (JJCIT), vol. 11, no. 2, pp. 211–237, 2025.

[11] X. Fu et al., "A Probabilistic Method for Image Enhancement with Simultaneous Illumination and Reflectance Estimation," IEEE Trans. Image Process., vol. 24, no. 12, pp. 4965–4977, 2015.

[12] X. Fu et al., "A Fusion-based Enhancing Method for Weakly Illuminated Images," Signal Process., vol. 129, pp. 82–96, 2016.

[13] X. Guo, Y. Li and H. Ling, "LIME: Low-light Image Enhancement *via* Illumination Map Estimation," IEEE Trans. Image Process., vol. 26, no. 2, pp. 982–993, 2017.

[14] Y. Ren, Z. Ying, T. H. Li and G. Li, "LECARM: Low-light Image Enhancement Using the Camera Response Model," IEEE Trans. Circuits Syst. Video Technol., vol. 29, no. 4, pp. 968–981, 2018.

[15] M. Tanaka, T. Shibata and M. Okutomi, "Gradient-based Low-light Image Enhancement," Proc. of the 2019 IEEE Int. Conf. on Consumer Electronics (ICCE), DOI: 10.1109/ICCE.2019.8662059, Las Vegas, NV, USA, Jan. 2019.

[16] J. Xie et al., "Semantically-guided Low-light Image Enhancement," Pattern Recognition Letters, vol. 138, pp. 308–314, 2020.

[17] N. Singh and A. K. Bhandari, "Principal Component Analysis-based Low-light Image Enhancement Using Reflection Model," IEEE Trans. Instrum. Meas., vol. 70, pp. 1–10, 2021.

[18] J. J. Jeon and I. K. Eom, "Low-light Image Enhancement Using Inverted Image Normalized by Atmospheric Light," Signal Process., vol. 196, p. 108523, 2022.

[19] Y. Demir and N. H. Kaplan, "Low-light Image Enhancement Based on Sharpening-Smoothing Image Filter," Digital Signal Processing, vol. 138, p. 104054, 2023.

[20] M. F. Hassan, T. Adam, H. Rajagopal and R. Paramesran, "A Hue Preserving Uniform Illumination Image Enhancement *via* Triangle Similarity Criterion in HSI Color Space," Visual Computer, vol. 39, no. 12, pp. 6755–6766, 2023.

[21] L. Wang, L. Zhao, T. Zhong and C. Wu, "Low-light Image Enhancement Using Generative Adversarial Networks," Scientific Reports, vol. 14, no. 1, p. 18489, 2024.

[22] I. M. Majid Mohammed and N. A. Mat Isa, "Contrast Limited Adaptive Local Histogram Equalization Method for Poor Contrast Image Enhancement," IEEE Access, vol. 13, pp. 62600–62632, 2025.

[23] S. Yang, D. Zhou, J. Cao and Y. Guo, "LightingNet: An Integrated Learning Method for Low-light Image Enhancement," IEEE Trans. Comput. Imaging, vol. 9, pp. 29–42, 2023.

[24] C. Zhang, K. M. Lam and Q. Wang, "CoF-Net: A Progressive Coarse-to-fine Framework for Object Detection in Remote-sensing Imagery," IEEE Trans. Geosci. Remote Sens., vol. 61, pp. 1–17, 2023.

[25] S. J. Im, C. Yun, S. J. Lee and K. R. Park, "Artificial Intelligence-based Low-light Marine Image Enhancement for Semantic Segmentation in Edge-intelligence-empowered Internet of Things Environment," IEEE Internet Things J., vol. 12, no. 4, pp. 4086–4114, 2025.

[26] C. Li, J. Guo, F. Porikli and Y. Pang, "LightenNet: A Convolutional Neural Network for Weakly Illuminated Image Enhancement," Pattern Recognition Letters, vol. 104, pp. 15–22, 2018.

[27] M. Afifi et al., "CIE XYZ Net: Unprocessing Images for Low-level Computer Vision Tasks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 44, no. 9, pp. 4688–4700, 2022.

[28] Y. Cai, H. Bian, J. Lin, H. Wang, R. Timofte and Y. Zhang, "Retinexformer: One-stage Retinex-based Transformer for Low-light Image Enhancement," Proc. of the IEEE/CVF Int. Conf. on Computer Vision (ICCV), pp. 12504–12513, 2023.

[29] Y. Cui, W. Ren, X. Cao and A. Knoll, "Revitalizing Convolutional Network for Image Restoration," IEEE Trans. Pattern Anal. Mach. Intell., vol. 46, no. 12, pp. 9423–9438, 2024.

[30] L. Xu, C. Hu, Y. Hu, X. Jing, Z. Cai and X. Lu, "UPT-Flow: Multi-scale Transformer-guided Normalizing Flow for Low-light Image Enhancement," Pattern Recognition, vol. 158, p. 111076, 2025.

[31] S. Bansal, R. K. Bansal and R. Bhardwaj, "A Novel Low Complexity Retinex-based Algorithm for Enhancing Low-light images," Multimedia Tools Appl., vol. 83, no. 10, pp. 29485–29504, 2024.

[32] A. Łoza et al., "Automatic Contrast Enhancement of Low-light Images Based on Local Statistics of Wavelet Coefficients," Digital Signal Processing, vol. 23, no. 6, pp. 1856–1866, 2013.

[33] M. Jourlin and J. C. Pinoli, "A Model for Logarithmic Image Processing," Journal of Microscopy, vol. 149, no. 1, pp. 21–35, 1988.

[34] X. Pei et al., "Robustness of Machine Learning to Color, Size Change, Normalization and Image Enhancement on Micrograph Datasets with Large Sample Differences," Materials & Design, vol. 232, p. 112086, 2023.

[35] O. Bryan et al., "A Diffusion-based Super Resolution Model for Enhancing Sonar Images," Journal of the Acoustical Society of America, vol. 157, no. 1, pp. 509–518, 2025.

[36] M. Ambrosanio, B. Kanoun and F. Baselice, "WKSR-NLM: An Ultrasound Despeckling Filter Based on Patch Ratio and Statistical Similarity," IEEE Access, vol. 8, pp. 150773–150783, 2020.
[37] S. Wang, J. Zheng, H. M. Hu and B. Li, "Naturalness Preserved Enhancement Algorithm for Non-uniform Illumination images," IEEE Trans. Image Process., vol. 22, no. 9, pp. 3538–3548, 2013.
[38] X. Min, G. Zhai, K. Gu, Y. Liu and X. Yang, "Blind Image Quality Estimation *via* Distortion Aggravation," IEEE Trans. on Broadcasting, vol. 64, no. 2, pp. 508–517, 2018.
[39] C. Gao, K. Panetta and S. Agaian, "Color Image Attribute and Quality Measurements," Proc. SPIE Mobile Multimedia/Image Processing, Security and Applications, vol. 9120, pp. 238–251, May 2014.
[40] N. Venkatanath, D. Praneeth, S. C. Sumohana and S. M. Swarup, "Blind Image Quality Evaluation Using Perception-based Features," Proc. of the 2015 21st National Conf. on Communications (NCC), pp. 1–6, Mumbai, India, 2015.
[41] W. Xue et al., "Blind Image Quality Assessment Using Joint Statistics of Gradient Magnitude and Laplacian Features," IEEE Trans. Image Process., vol. 23, no. 11, pp. 4850–4862, 2014.
[42] N. Singh and A. K. Bhandari, "Noise Aware L_2–LP Decomposition-based Enhancement in Extremely Low Light Conditions with Web Application," IEEE Trans. on Consumer Electronics, vol. 68, no. 2, pp. 161–169, 2022.

ملخص البحث:

في بعض الأحيان، يتمّ التقاط الصُّور الأثرية في بيئات تتّسم بإضاءةٍ غير مثالية، الأمر الّذي يؤدّي إلى إضاءةٍ غير متوازنة وفقدان بعض التّفاصيل. وتعملُ مثل هذه المشكلات على إعاقة بعض العمليات، مثل التّحليل والتّفسير والتّمثيل والنّمذجة ثلاثية الأبعاد.

تُقدّم هذه الدّراسة خوارزميةً من شأنها أن تعمل على تحقيق توازن الإضاءة، وهي خوارزمية لوغاريتمية-إحصائية تحمِلُ اسم "Log-Stat"، وتتكوّن من مرحلتين: الأولى تَستخدم ستَّ عملياتٍ رياضية، بينما تستفيد الثانية من أربعِ عملياتٍ.

تمّ استخدامُ صُور متنوّعة لفحص الخوارزمية، وجرت مقارنة الخوارزمية المقترحة مع عَشر خوارزميات مشهورة، وذلك باستخدام ستّة مؤشّرات أداءٍ. وقدْ برهنت المقارنة على نجاح الخوارزمية المقترحة في عددٍ من الجوانب المختلفة، منها استعادة الوضوح وتحقيقُ توازن الإضاءة، الأمر الّذي يؤدّي إلى دقّة في التّفاصيل الّتي يمكن أن تتأثّر سلباً بغياب توازن الإضاءة في الصُّور الأصلية. والجدير بالذّكر أنّ ذلك يؤكّد نجاعة الخوارزمية المقترحة في هذه الدّراسة في تحسين جودة الصُّور الأثرية الملتقطة في ظروف إضاءةٍ غير مثالية.

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 12, No. 01, March 2026.

MULTI-CLASS HEART DISEASE CLASSIFICATION USING MULTI-LEAD ECG FEATURES AND ENSEMBLE LEARNING

Bala Venkateswarlu Isunuri, Venkata Sravya Alapati, Sri Charan Marripudi, Gopala Krishna Parimi and Akshaya Valli Koganti

(Received: 28-Sep.-2025, Revised: 13-Dec.-2025, Accepted: 6-Jan.-2026)

ABSTRACT

Cardiovascular diseases (CVDs) are the leading causes of global mortality and require an early and precise diagnosis. This work presents an automated multi-class classifier for diagnosing cardiac disease from electrocardiogram (ECG) images through image processing and machine-learning techniques. The proposed framework consists of three steps, including pre-processing, feature extraction, and ensemble learning. Initially, the ECG image undergoes a comprehensive pre-processing pipeline that includes lead segmentation, grayscale conversion, Gaussian filtering, and Otsu thresholding. The contour-based features are extracted and then reduced by PCA to preserve discriminative information. Finally, multiple machine-learning models, including K-nearest neighbors (KNNs), Random Forest and support vector machines (SVMs), are ensembled using voting and stacking classifiers to improve the performance of the proposed framework. The proposed ensemble model is evaluated on a public dataset that consists of ECG images that are categorized into four classes: normal, abnormal, myocardial infarction (MI), and history of MI. The proposed ensemble model attained the highest classification accuracy of 98.06% and outperformed the existing pre-trained and state-of-the-art models.

KEYWORDS

Electrocardiogram, Ensemble learning, Multi-lead ECG features, Multi-class heart disease classification.

1. INTRODUCTION

Cardiovascular diseases (CVDs) are a major public-health issue and the leading cause of death worldwide. The World Health Organization (WHO) says that CVDs kill 17.9 million people each year, which is 31% of all deaths in the world [2]. CVDs are a category of diseases that include coronary artery disease, myocardial infarction (MI), heart failure, and arrhythmias. These diseases don't show up until later stages. For early medical treatment, lower healthcare costs, and a better quality of life, it is important to diagnose heart disorders correctly and early. Electrocardiography (ECG) is the most relevant and extensively employed non-invasive diagnostic modality for evaluating cardiac electrical activity. A conventional ECG provides important information about the heart's rhythm, conduction patterns, and how anomalies show up, which helps doctors figure out what kind of heart disease a person has [3]. The physical examination of an ECG incurs considerable time and practice, and the results may vary among individuals. This constraint makes it exceedingly challenging to identify by a cardiologist in rural or under-resourced regions. This issue has generated a demand for automated ECG analysis solutions utilizing artificial intelligence (AI) to enhance diagnostic accuracy and scalability for physicians. Recent breakthroughs in machine learning (ML) and deep learning (DL) have significantly revolutionized the domain of biological signal processing. Techniques, such as convolutional neural networks (CNN), recurrent neural networks (RNNs) and ensemble-learning models, have achieved unprecedented advancements in the classification of cardiac diseases utilizing ECG waveforms and pictures [4]-[5]. The models can discover valuable traits, discern subtle trends, and manage extensive data volumes with limited human involvement.

CNN-based structures have widely been applied to image-related ECG classification applications, as they have superior capabilities in spatial-learning features [6]. Some previous studies have successfully applied these methods, yielding reasonable outcomes. In [1], the authors employed CNN models like MobileNetV2 and VGG16, for four-class classification of ECG images into Normal, Abnormal, Myocardial Infarction (MI), and History of MI. The method employed real-time deployment by

---

B. V. Isunuri (Corresponding Author), V. S. Alapati, S. C. Marripudi, G. K. Parimi and A. V. Koganti are with the Department of Computer Science and Engineering, SRM, University AP, Amaravati, Andhra Pradesh, India. Emails: {`bala.v`, `venkatasravya_a`, `sricharan_m`, `gopalakrishna_p` and `akshayavalli_k}@sramp.edu.in`

executing the model on Raspberry Pi platforms and obtained classification accuracy rates of up to 95%. The authors largely trained the model using end-to-end learning without proper signal pre-processing and lead-wise segmentation, potentially limiting its interpretability and adaptability. Other works followed the traditional machine-learning approach by employing hand-designed pre-processing pipelines with ML classifiers. For instance, works in [9] and [10] pre-processed ECG images by segmenting them into 12 leads, computed statistics and shape-based features, and employed classifiers, like Support Vector Machines (SVMs), K-Nearest Neighbors (KNNs), and Logistic Regression. Such pipelines employed Principal-component Analysis (PCA) as a feature reducer and utilized voting classifiers for class accuracy improvement, with the best accuracy of 94.2%. They employed only 12 leads and did not follow strict validation regimes, like K-fold cross-validation. Hence, they are less robust and less generalizable.

We present a new and complete ECG-based cardiac-disease classification system that takes the best of these previous works and extends them. Our contribution is 13 -lead ECG segmentation that records a wider and more nuanced spatial coverage of the heart's electrical activity than conventional 12-lead systems. Each lead is treated separately, enabling our system to learn local features from alternative heart views. This approach alone makes our model structurally and physiologically different, enabling better representation of subtle ECG waveform abnormalities. Further, our ensemble technique aggregates the predictions of multiple base classifiers, including SVMs, KNNs, logistic regression, and XGBoost, to achieve maximum diversity and performance. The contributions of this paper are multifold:

1. Adding 13-lead segmentation to improve spatial resolution.
2. A pre-processing pipeline with grayscale, Gaussian filtering, and Otsu thresholding.
3. The system integrates multiple conventional classifiers using ensemble-learning methods.
4. Cross-validation is utilized to obtain a precise performance estimate.

These contributions together offer a scalable, interpretable, and accurate solution for automated ECG interpretation with high potential for application in real-world clinical decision-support systems and portable diagnostic devices. The paper is organized as follows: Section 1 introduces multi-class cardiovascular-disease classification. The literature review is presented in Section 2. Details of the proposed ensemble framework are explored in Section 3. Section 4 includes an analysis of the results and a discussion, while Section 5 concludes the findings.

2. LITERATURE REVIEW

In recent years, the application of artificial-intelligence (AI) methods in automated electrocardiogram (ECG) data processing for the identification and classification of cardiovascular diseases has significantly increased. Conventional methods rely heavily on clinical expertise and human judgment, resulting in subjectivity or delays. Researchers have sought to utilize both conventional machine-learning (ML) and deep-learning (DL) techniques to improve diagnostic accuracy and scalability. Various studies have examined deep-learning models for the classification of ECG images. Lightweight CNN architectures, such as MobileNetV2 and VGG16, have been employed to extract discriminative features from ECG images, facilitating real-time implementation on embedded systems [1]. The models demonstrated consistent effectiveness in detecting cardiac anomalies, including myocardial infarction and arrhythmias, with an accuracy of up to 95%. Convolutional Neural Networks (CNNs) require extensive datasets and GPU-based training and may lack interpretability-factors that can hinder their clinical implementation in resource-constrained settings [3]-[4]. Nevertheless, conventional machine-learning techniques have gained prominence due to their cost-effectiveness and comprehensibility. Effective approaches for ECG classification include support vector machines (SVMs), K-Nearest Neighbors (KNNs), logistic regression, and XGBoost. These methods are accurate when utilized alongside well-crafted features [2][5][7]. Although the models demonstrate robust performance, they may exhibit heightened sensitivity to noise and extraneous features, particularly when handling high-dimensional data. Ensemble approaches, like voting and stacking, have been proposed to tackle these issues. These are predicated on the robustness of a mixture of multiple base models to enhance their durability and generalization capabilities. Voting classifiers employ majority voting to combine the judgments, whereas stacking uses a meta-model to select the ideal combination of basic outputs to reach improved accuracy [6], [9]. For example, [9] employed ensemble models optimized by GridSearch, attaining a peak accuracy of 92.4%. Pre-processing is an essential component in rendering feature extraction significant. Gray-scale conversion, Gaussian filtering, and Otsu's thresholding are methods commonly employed to enhance the clarity of the ECG waveform [2][5][10]. Principal-component Analysis (PCA) is employed to diminish feature dimensionality while preserving the most pertinent information with decreased computational effort. The latest research is confined to standard 12-lead ECG pictures.

Finally, the literature encompasses a broad spectrum of CVD detection from ECGs. However, there is still a gap in performance optimization that does not compromise interpretability or increase computational complexity. Our contribution bridges these gaps by putting forward an ensemble learning-based classification pipeline with improved performance without deep architectures. In the proposed framework, we incorporate 13-lead segmentation, providing greater spatial resolution of the cardiac activity. We incorporate ensemble approaches, including voting and stacking with cross-validation, for robust and balanced assessment. These additions make our model more capable of performing better than conventional ML pipelines [9]-[10] as well as deep CNN-based models, like MobileNetV2 [1].

3. METHODOLOGY

Let $X = \{x_1, x_2, \dots, x_n\}$ be the set of input ECG images, where every $x_i \in \mathbb{R}^{2213\times1572\times3}$. Similarly, $Y = \{y_1, y_2, \dots, y_n\}$ represents class labels of each x_i, where $y_i \in \{0,1,2,3\}$. Table 1 lists the class description for each class. The proposed framework comprises three stages for multi-class heart-disease classification. It accepts the input ECG image x_i and produces the class label y_i, as shown in Figure 1. The details of each stage are as follows.

Table 1. Description of each class label.

Class label	Description
0	HB (Abnormal Heart Beat)
1	MI (Myocardial Infraction)
2	Normal
3	HMI (History of MI)

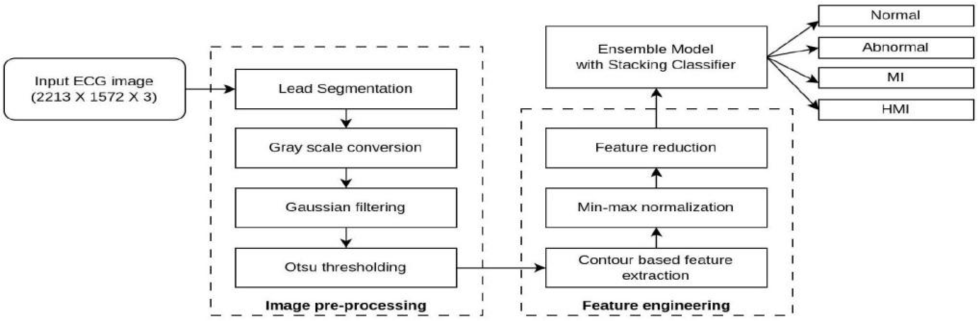


Figure 1. The proposed multi-lead features and ensemble learning framework.

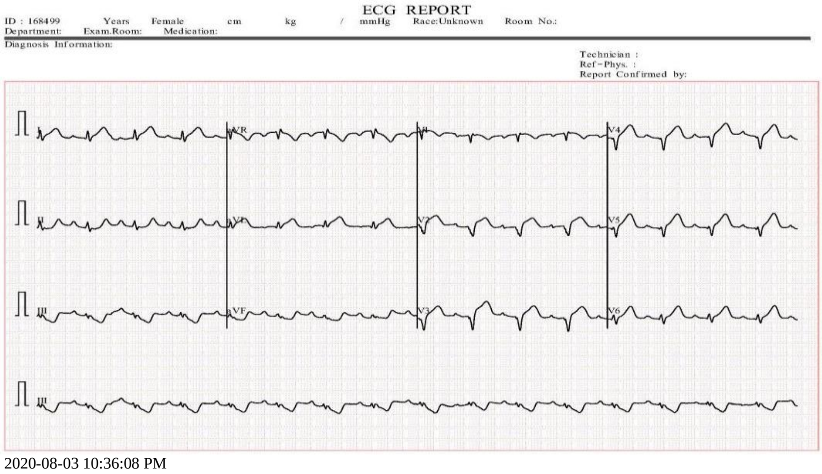


Figure 2. Sample ECG image with 13 leads.

3.1 Image Pre-processing

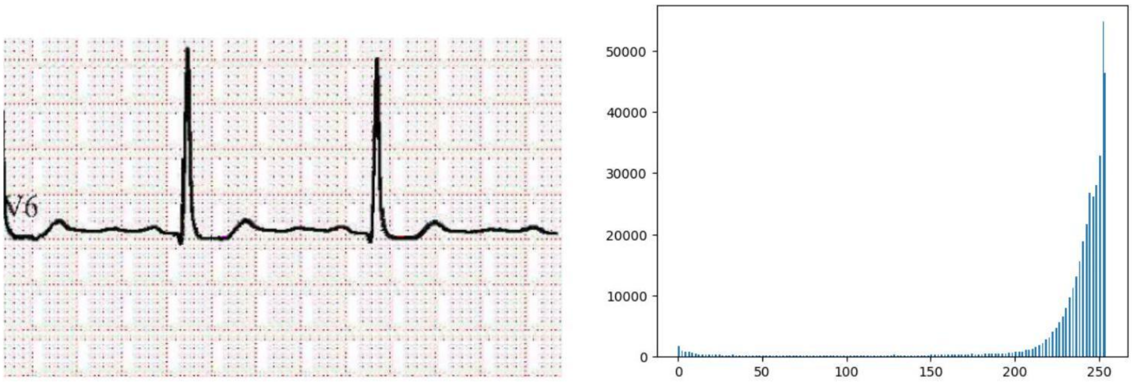
The proposed framework utilizes a specialized pre-processing pipeline to enhance ECG-image features. The pre-processing of images involves four steps, and the details are as follows.

1) Lead Segmentation

The literature reveals that existing models utilize 12 ECG leads for the classification of heart diseases [22]. Madias [19] presented the importance of lead 13 of ECG in heart diseases. This motivated us to involve the 13th lead for the proposed framework. Thus, each ECG image x_i is divided into 13 independent leads, each of which corresponds to different electrical activities captured from different parts of the heart. The dataset shows that each lead L_j has fixed coordinates. Thus, coordinate slicing across standard spatial positions for all images was used to mark the 13 leads $L = \{L_1, L_2, \dots L_{13}\}$, where every L_j is a single lead image. Figure 2 represents a sample ECG image acquired from the dataset [20]. The leads are obtained from three horizontal bands and one bottom strip. Leads 1-4 are taken from the top section (rows 300-600), leads 5-8 from the middle band (rows $600 - 900$), and leads $9 - 12$ from the bottom band (rows $900 - 1200$). Lead 13 spans rows 1250-1480 and spans the entire horizontal range, which is known as an extended rhythm strip. This partitioning allows for uniform spatial extraction of waveforms for all samples. The derived leads are harvested based on actual coordinate regions as follows:

$$\begin{gathered}
L_1 \leftarrow x_i[300{:}\,600{,}150{:}\,643] \\
L_2 \leftarrow x_i[300{:}\,600{,}646{:}\,1135] \\
L_3 \leftarrow x_i[300{:}\,600{,}1140{:}\,1625] \\
L_4 \leftarrow x_i[300{:}\,600{,}1630{:}\,2125] \\
L_5 \leftarrow x_i[600{:}\,900{,}150{:}\,643] \\
L_6 \leftarrow x_i[600{:}\,900{,}646{:}\,1135] \\
L_7 \leftarrow x_i[600{:}\,900{,}1140{:}\,1625] \\
L_8 \leftarrow x_i[600{:}\,900{,}1630{:}\,2125] \\
L_9 \leftarrow x_i[900{:}\,1200{,}150{:}\,643] \\
L_{10} \leftarrow x_i[900{:}\,1200{,}646{:}\,1135] \\
L_{11} \leftarrow x_i[900{:}\,1200{,}1140{:}\,1625] \\
L_{12} \leftarrow x_i[900{:}\,1200{,}1630{:}\,2125] \\
L_{13} \leftarrow x_i[1250{:}\,1480{,}150{:}\,2125]
\end{gathered} \tag{1}$$

In general, ECG images consist of red-colored grid-like patterns on which black-colored ECG signals are printed. Figure 3(a) and Figure 3(b) depict the 12th lead segment color image and corresponding histogram of the sample ECG image shown in Figure 2. From the histogram, it can be observed that there is a huge number of pixels (around 50000) with high intensity above 250. Simple thresholding will cause loss of ECG signal information. Thus, the proposed framework utilizes a sequence of pre-processing steps to preserve the ECG wave pattern.



(a) (b)

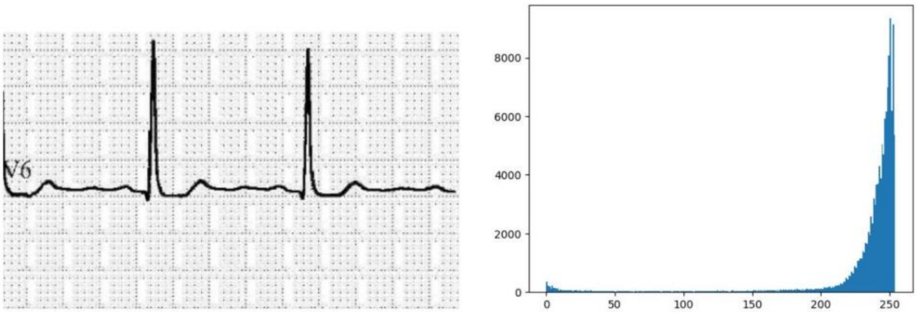
Figure 3. Color image of 12th lead segment along with its histogram.

2) Grayscale Conversion

Each color lead image (L_j) is transformed to grayscale to reduce dimensionality and emphasize signal morphology over color. The grayscale conversion is defined by Equation 2.

$$L_{\text{gray}}(x,y) = 0.29 \cdot R(x,y) + 0.58 \cdot G(x,y) + 0.11 \cdot B(x,y) \tag{2}$$

The red, green, and blue intensities of pixel (x, y) are represented by R, G, and B . The value of $L_{\text{gray}}(x,y)$ represents the resulting grayscale intensity for pixel (x, y). This conversion maximizes downstream processing and excludes color data that is irrelevant to the analysis of ECG waveforms. Figure 4(a) and Figure 4(b) depict grayscale of 12th lead segment color image and corresponding histogram, respectively. From the histogram, it can be observed that the peak number of pixels has reduced to 10000 due to single channel.



(a) (b)

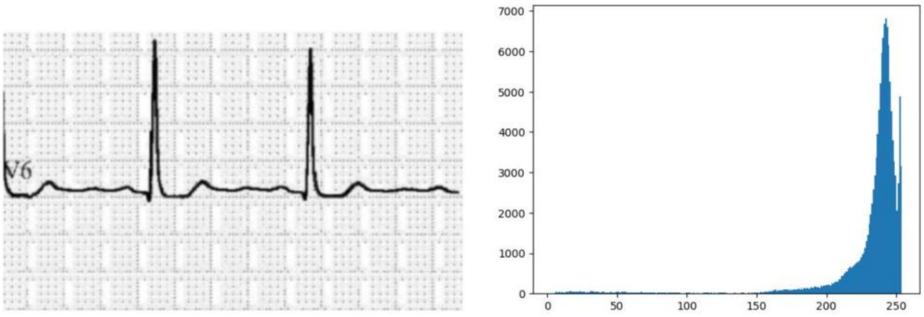
Figure 4. Grayscale image of the 12th lead segment along with its histogram.

3) Gaussian Filtering

The grayscale image was filtered with a Gaussian filter to cut down on high-frequency noise and blur it without losing the structural outlines [21]. The two-dimensional Gaussian kernel looks as follows:

$$G(x,y) = \frac{1}{2\pi\sigma^2} \cdot \exp{}^{-\left(\frac{x^2+y^2}{2\sigma^2}\right)} \tag{3}$$

In this case, σ stands for the standard deviation of the Gaussian filter. It tells the quantity of smoothing, and the blurring effect of the Gaussian filter can be controlled using the σ value. The ECG images need more smoothing to suppress the effect of the background grid pattern. Thus, we have considered high sigma as $\sigma = 0.7$. Similarly, kernel size is another parameter that influences the feature quality. Thus, we have conducted an ablation study with different kernel sizes to identify a better kernel size. The results of the ablation study are reported in the results section. The filtered image is the result of the convolution of the grayscale image with the Gaussian kernel generated from Equation 3. The process smooths the ECG trace lines by eliminating background variance and minor graphical artifacts. Figure 5(a) and Figure 5(b) depict the smoothened image of the 12th lead segment image and the corresponding histogram, respectively. From the histogram, it can be observed that the peak number of pixels has further reduced to 7000 due to suppression of background pixels.



(a) (b)

Figure 5. Smoothened image of the 12th lead segment along with its histogram.

4) Otsu's Thresholding

We apply Otsu's method to turn the smoothened image into binary and make the ECG signal stand out as a binary image. Otsu's thresholding finds the best threshold that maximizes the difference between background and foreground pixels [21]. The optimal threshold t is the one that maximizes $\sigma_b^2(t)$.

$$\sigma_b^2(t) = \omega_0(t) \cdot \omega_1(t) \cdot [\mu_0(t) - \mu_1(t)]^2 \tag{4}$$

In this context, the symbols $\omega_0(t)$ and $\omega_1(t)$ represent the probabilities (weights) associated with the background and foreground classes, respectively. Similarly, $\mu_0(t)$ and $\mu_1(t)$ are the mean intensities of background and foreground classes. $\sigma_b^2(t)$ is between-class variance at threshold t. Finally, the Otsu thresholded binarized image is resized to 300×450 to reduce computational complexity. Figure 6 depicts the Otsu thresholded 12th lead segment, which removes the grid pattern of the ECG color image.

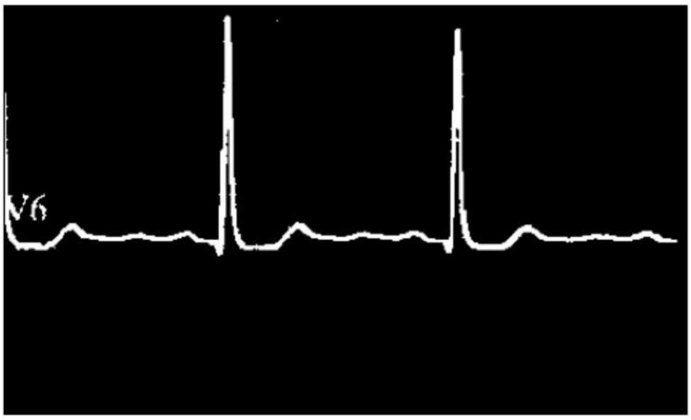


Figure 6. Otsu thresholded binarized image of 12th lead segment.

This phase transforms input image x_i to 13 binarized lead images $BL = \{BL_1, BL_2, \dots BL_{13}\}$ with improved contrast between the ECG waveform and background. This pre-processing sequence guarantees that the signal contours are maintained and evidently separable, allowing for efficient feature extraction in the subsequent phase.

3.2 Feature Engineering

Contour-based features are employed in the proposed framework. Therefore, the feature-engineering process consists of three sub-tasks as outlined below.

1) Contour-based Feature Extraction

After Otsu thresholding, contour detection is used to identify contours and boundaries of the shapes within the segmented leads. From the contours, morphological characteristics are obtained to describe significant structural characteristics of the segmented leads. The characteristics may be shape, size, and orientation. Other morphological features may include perimeter, aspect ratio, convexity, …etc. Figure 7 depicts the contour plot of the 13th lead segment. From that contour plot, only 255 points from each lead will be selected at equal internals.

2) Min-Max Normalization

The range of contour-based features depends on the lead number, which leads to varying scales and distributions. Therefore, normalization of the features extracted from each lead is necessary. We have considered min-max normalization to normalize the features using the following equation.

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{5}$$

where X is the original feature value. $X_{\min}$ and $X_{\max}$ are the lower and upper limits of the feature, respectively. X' is the normalized feature value; now the values of each lead image are in the range $[0,1]$. After normalization, each lead vector has 255 values. Thus, this stage generates a feature vector F with $13 \times 255 = 3315$ features for each ECG image.

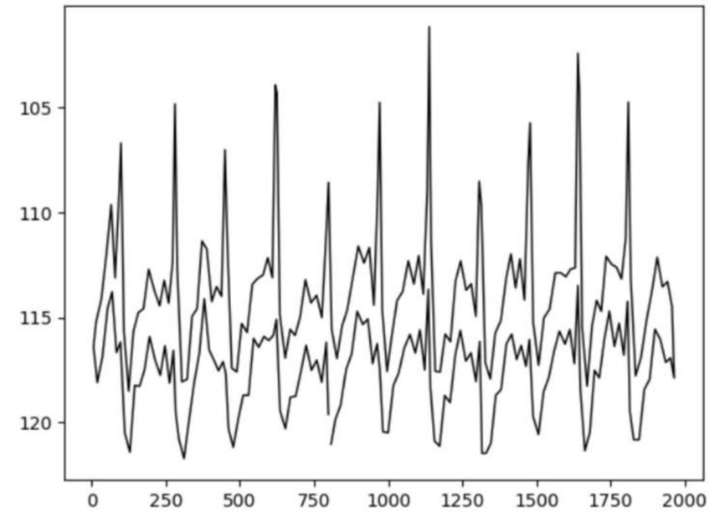


Figure 7. Contour plot of 13th lead segment.

3) Feature Reduction

To make ensemble learning less complicated, we used feature reduction on the feature vector F obtained from the previous phase. Principal-component Analysis (PCA) is a way to reduce the number of dimensions of features by mapping them onto a new collection of orthogonal components (principal components). PCA seeks to get rid of redundant information while keeping the most useful parts. The PCA implementation finds the eigenvalues and eigenvectors of the covariance matrix and uses them to move the data into a new sub-space, which makes it smaller. The output of the previous step produces a huge number of features (3315 features). It needs a feature reduction to optimize the computational cost of the proposed model. However, the number of features is significant for how well ensemble learning works. Therefore, we conducted an ablation study with varying numbers of features to achieve the optimal performance. We discovered in our ablation study that a feature vector with 512 features performs the best of the options that we looked at. This stage generates a feature vector with 512 normalized contour-based features.

3.3 Ensemble Learning

We used machine learning algorithms, including Support Vector Machines (SVMs), K-Nearest Neighbors (KNNs), and Logistic Regression (LR) to create the proposed ensemble framework. SVM builds a maximum-margin hyperplane to separate classes in a high-dimensional space. SVM is able to handle nonlinear data by employing kernel tricks and generalizes extremely well. SVM is particularly effective when applied to small- to medium-sized biomedical datasets. KNN is a very basic, non-parametric classifier that computes the majority vote of the K-nearest neighbors. Relevant features and high-dimensional data adversely affect KNN's ability to maintain local patterns. LR is class membership probability as a logistic function. LR is computationally efficient, interpretable, and a useful linear baseline for multi-class classification using the one-vs.-rest strategy. XGBoost is an outstanding ensemble method that is gradient-boosted decision tree-based. XGBoost performs best with complex patterns, restricts overfitting, and offers high accuracy as well as efficient operation. The individual training and testing of all models on the PCA-reduced feature vectors are performed to create the performance baselines. Two ensemble techniques were employed to obtain the highest individual model and lowest predictive consistency.

- Voting classifier: This ensemble method uses soft voting to combine the probability output of chosen base classifiers. Classification was performed by averaging predicted probabilities and selecting the best combined score label.
- Stacking classifier: The base-level classifiers' predictions are used as input to a meta-classifier.

The second-level learner learns the base models' dependencies, which enabled improved decision boundaries and classification outcomes.

Ensemble models are created to take advantage of the complementary strengths of each algorithm and to boost the overall robustness. The proposed ensemble model utilized the grid search-based cross validation method with five folds for hyper-parameter tuning. We achieved the best performance with an ensemble model that consists of SVM with $C = 1$, SVM with Gamma=0.01, KNN with 5 neighbors and Random Forest with 300 trees.

4. RESULTS

This section presents a detailed performance analysis of the proposed ensemble model along with an ablation study. Model performance was approximated using common classification metrics, including accuracy, precision, recall, and F1-score. The proposed model has been evaluated on the Mendeley Dataset [20]. The dataset holds 928 ECG color images with a resolution of 2213×1572. Each image belongs to one of the four cardiac conditions: normal, abnormal, myocardial infarction (MI), and history of MI (HMI).

4.1 Ablation Study

Gaussian filtering and feature reduction primarily influence the performance of the proposed model. Therefore, we conducted an ablation study focusing on these two aspects.

- **With different Gaussian filter sizes:**

  Firstly, we have considered 100 features and experimented with different Gaussian filter sizes in the pre-processing step, including $3 \times 3, 5 \times 5$ and 7×7. Table 2 represents a comparison of performance with different Gaussian filter sizes. This table reveals that the logistic-regression model exhibits the least performance, and the XGBoost model attains the highest performance among the basic machine-learning models. However, the proposed stacking-based ensemble model outperforms basic models with $94.5\%, 96.2\%$, and 93.4% with $3 \times 3, 5 \times 5$, and 7×7 filter sizes, respectively. It also indicates that the proposed stacking-based ensemble model degrades its performance with a 7×7 filter size. The proposed stacking classifier exhibits competing performance with Gaussian filter having 3×3 and 5×5 filter sizes. However, the proposed model with Gaussian filter having 5×5 filter size exhibits best performance.

Table 2. Ablation study with different Gaussian filter sizes.

	Filter size		
Model	3×3	5×5	$\mathbf{7} \times 7$
KNN	80.5	84.8	81.7
Logistic Regression	76.6	79.3	81.1
SVM	91.2	92.9	91.5
XGBoost	92.2	93.5	91.8
Voting Classifier	91.3	94.2	92.0
Stacking Classifier	$\mathbf{94.5}$	$\mathbf{96.2}$	$\mathbf{93.4}$

- **With different number of features:**

  In general, the number of features plays a vital role in the performance of a model. Thus, we experimented with different numbers of features, including 100, 400, and 512, in the feature-reduction process. We analyzed the proposed model with Gaussian filter sizes 5×5 and 3×3 to understand how it behaves. Table 3 represents a performance comparison with different numbers of features having a Gaussian filter size of 5×5. In this case, the proposed stacking-based ensemble model exhibits the best performance with 400 features. Similarly, Table 4 represents a performance comparison with different numbers of features having a Gaussian filter size of 3 × 3. In this case, the proposed ensemble model exhibits the best performance at 98% with 400 and 512 features. Tables 3 and 4 reveal that the proposed stacking-based ensemble model exhibits the best performance with a Gaussian filter size of 3 × 3 with 400 and 512 features. The computational cost of the proposed model with 400 features is less than that with 512 features. Further, the proposed model exhibits similar performance in both cases. Hence, we have considered 400 features to optimize the proposed stacking-based ensemble model.

Table 3. Ablation study with different numbers of features having Gaussian filter size of 5 × 5.

	Number of features		
Model	**100**	**400**	**512**
KNN	84.8	78.4	77.2
Logistic Regression	79.3	89.3	89.6
SVM	92.9	94.2	94.1
XGBoost	93.5	93.2	96.3
Voting Classifier	94.2	97.3	**97.7**
Stacking Classifier	**96.2**	**97.7**	95.7

Table 4. Ablation study with different numbers of features having a Gaussian filter size of 3 × 3.

	Number of features		
Model	**100**	**400**	**512**
KNN	80.5	77.4	76.8
Logistic Regression	76.6	89.0	88.9
SVM	91.2	93.0	93.3
XGBoost	92.2	93.4	97.1
Voting Classifier	91.3	95.9	96.7
Stacking Classifier	**94.5**	**98.1**	**98.3**

4.2 Performance Comparison of Proposed Ensemble Model

The existing models utilized only 12 leads for the ECG classification. For more in-depth examination and to mimic clinical lead-based interpretation, ECG images were divided into 13 leads rather than the traditional 12-lead setup. We have considered popular machine-learning models for the performance analysis. In the proposed model, we utilized a Gaussian filter with a 3 × 3 kernel filter size in pre-processing and PCA with 400 features for feature reduction. The proposed ensemble model also consists of an ensemble of SVM with $C = 1$, SVM with Gamma $= 0.01$, KNN with 5 neighbors, and Random Forest with 300 trees. Table 5 lists out class-wise performance of proposed stacking classifier.

Table 5. Class-wise performance of proposed stacking classifier.

Class	**Precision**	**Recall**	**F1-Score**
0	100.00	100.00	100.00
1	100.00	100.00	100.00
2	97.95	95.79	96.80
3	93.52	96.54	94.87
Accuracy	98.06		
Weighted Avg.	98.17	98.06	98.06

Table 6. Model performance comparison with ML models.

Model	Accuracy	Precision	Recall	F1-Score
KNN	77.36	77.79	77.37	75.94
Logistic Regression	88.69	89.47	88.69	88.19
SVM	93.10	93.27	93.10	93.02
XGBoost	93.32	93.56	93.32	93.24
Voting Classifier	95.47	95.85	95.47	95.36
Stacking Classifier	98.06	98.17	98.06	98.06

Table 6 lists out performance comparisons of stacking-based ensemble models with basic machine-learning models. Among the individual classifiers, SVM and XGBoost were the best performers, with over 93% accuracy, followed by logistic regression in the third place with 88.69%. KNN was the worst performer with lower metrics owing to its vulnerability to noisy and high-dimensional data. The SVM model gave consistent classification accuracy for each of the four classes of ECG with a total accuracy of 93.10%. Weighted precision, recall, and F1-score were 93.27%, 93.10%, and 93.02%, respectively. The XGBoost classifier achieved an accuracy of 93.32%, weighted precision of 93.56%, recall of 93.32%, and F1-score of 93.24%. Its gradient-boosting mechanism allowed it to handle non-linear decision boundaries well, since it could capture the fine changes of ECG waveform features. However, the proposed ensemble model performs better on all measures, with an accuracy rate of above 95%. The voting classifier achieved 95.47% accuracy that uses a soft vote algorithm on predictions from SVM, KNN and Random Forest. The stacking classifier, which uses a meta-model to integrate the predictions of different base learners, like SVM, KNN, and Random Forest, also achieved the greatest overall performance, with 98.06% accuracy and a balanced F1-score of 98.06%. Its hierarchical-learning framework enables the comprehension of intricate interactions across models and enhances predictive accuracy, rendering it the optimal selection for multi-class classification challenges. The data suggests that utilizing ensemble methods, especially stacking, improves the accuracy, reliability, and overall effectiveness of diagnosing cardiovascular problems through ECG.

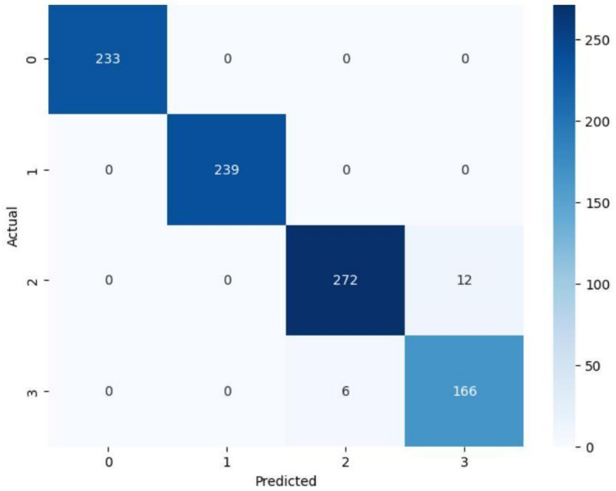


Figure 8. Confusion matrix of the proposed ensemble model.

The confusion matrix indicated that all four classes had more stability in their categorization. The ensemble mechanism provided the model with optimal confidence in recognizing borderline cases. The Abnormal and HMI classes exhibited a reduction in both false positives and false negatives. The results indicate that integrating various base learners enhances the system's resilience and reliability. Figure 8 illustrates the confusion matrix for the proposed stacking-based ensemble model. The confusion matrix indicates that the stacking classifier effectively distinguished between the classes, exhibiting minimal errors. The stacking classifier effectively distinguished between the MI and HMI classes, demonstrating exceptional precision and recall, indicative of robust discriminatory capability. The stacking design exhibited superior performance due to its ability to use inter-model interactions.

Table 7. Performance comparison with state-of-the-art models.

Model	**Accuracy (%)**
MobileNetV2 Transfer Learning [1]	93.00
MobileNetV2 Fine Tuning [1]	95.00
VGG16 Transfer Learning [1]	91.00
VGG16 Fine Tuning [1]	95.00
Sakli et al. [7]	96.70
Proposed Voting Classifier	95.47
Proposed Stacking Classifier	**98.06**

The efficacy of the proposed stack-based ensemble model is compared with that of existing models, as illustrated in Table 7. The findings indicate that the proposed stack-based ensemble model surpasses current state-of-the-art models, achieving an accuracy of 98.06%.

4.3 Discussion

This paper presents an extremely accurate model for predicting cardiovascular disease (CVD) based on ensemble machine-learning models. Of all the classifiers examined, the best performance was exhibited by the stacking classifier at 98.06% accuracy. Compared to [1], which used computationally costly models, like MobileNetV2 (94%) and VGG16 (92%), our stacking classifier outperformed them, even though it used less computationally costly models. This result proves the feasibility of ensemble learning even without using computationally costly models. Reference [7] employed a range of standard machine-learning models, including KNN, logistic regression, XGBoost, and SVM, and achieved a best accuracy of 96.7%. In [8], there were only regular ML techniques tried out, and the highest documented accuracy was less than 92.4%. Reference [9] used the models mentioned above, such as SVM, KNN, RF, and logistic regression, achieving an accuracy up to 92.4% with ensemble learning using GridSearch to optimize.

These comparisons also highlight the fact that, although there are excellent deep-learning architectures, such as MobileNetV2 and VGG16, well-hyperparameterized ensemble machine-learning algorithms can provide similar or even superior performance without the need for deep neural networks. This not only makes our method accurate, but also lightweight, interpretable, and computationally efficient with a significant advantage in real-world deployments to resource-constrained environments. The performance summary shows that ensemble-learning methods are better than regular classifiers when it comes to detecting heart diseases using ECGs.

5. CONCLUSION

This paper presents a scalable framework for multi-class classification of cardiovascular diseases from ECG images. ECG images can be processed for multi-class heart-disease classification through better pre-processing, contour-based feature extraction, and an ensemble-learning pipeline. Our results indicate that the ensemble-stacking classifier significantly outperforms individual models and all the earlier published works. The stacking classifier, with an accuracy of 98.06%, not only performed better than traditional machine-learning models, but also deep learning-based classifiers, such as MobileNetV2 and VGG16. Compared with deep architectures, the improved performance and reduced computational load of our architecture render it highly suitable for real-world use in resource-constrained environments. Additional class-wise accuracy improvement, particularly for valuable classes, like MI and HMI, also renders the system more practical. Briefly, the work illustrates the ability of carefully crafted traditional and ensemble-learning approaches to state-of-the-art performance on the cardiac-disease detection from ECG images. Some potential areas of future extensions of this paper include its application in real time on edge hardware, integration in clinical decision-support systems, or multi-modal health data to more general diagnostic applications.

REFERENCES

[1] L. Mhamdi, O. Dammak, F. Cottin and I. B. Dhaou, "Artificial Intelligence for Cardiac Diseases' Diagnosis and Prediction Using ECG Images on Embedded Systems," Biomedicines, vol. 10, no. 8, 2022.

[2] M. T. Tahmid, M. E. Kader, T. Mahmud and S. A. Fattah, "MDCardioNet: A Multi-dimensional Deep Neural Network for Cardiovascular Disease Diagnosis from Electrocardiogram," IEEE Journal of Biomedical and Health Informatics, vol. 28, no. 4, pp. 2005-2013, April 2024.

[3] M. B. Abubaker and B. Babayigit, "Detection of Cardiovascular Diseases in ECG Images Using Machine Learning and Deep Learning Methods," IEEE Transactions on Artificial Intelligence, vol. 4, no.2, pp. 373-382, April 2023.

[4] N. Mitra and B. I. Morshed, "Analyzing Clinical 12-lead ECG Images Using Deep Learning Algorithms for Objective Detection of Cardiac Diseases," Proc. of the 2022 IEEE 13th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), pp. 0517-0523, New York, USA, 2022.

[5] M. Balipa, M. Murugappan and A. Castalino, "Artificial Intelligence-based ECG Analysis to Assess Cardiac Health," Proc. of the 2024 3rd Int. Conf. on Sentiment Analysis and Deep Learning (ICSADL), pp. 587-593, Bhimdatta, Nepal, 2024.

[6] M. Gulhane and S. Kumar, "Deep Learning based Heart Diseases Detection Using ResNet50 Architecture," Proc. of the IEEE 2023 Intelligent Methods, Systems and Applications (IMSA), pp. 193-198, Giza, Egypt, 2023.

[7] M.Sakli, N. Sakli and H. Sakli, "ECG Images Automated Diagnosis based on Machine Learning Algorithms," Proc. of the 2023 20th Int. Multi-Conf. on Systems, Signals & Devices (SSD), pp. 934-939, Mahdia, Tunisia, 2023.

[8] M. D. Gresa and V. Sathya, "Exploring the Potential of Machine Learning and Deep Learning in ECG Image Analysis for Cardiovascular Disease Diagnosis," Proc. of the 2024 5th Int. Conf. on Electronics and Sustainable Communication Systems (ICESC), pp. 13441349, Coimbatore, India, 2024.

[9] D. Rautela, M. Bajeli, A. Kumar and H. Vaidya, "Identifying Cardiovascular Disorders through ECG Image Analysis," Proc. of the 2024 Int. Conf. on Cognitive Robotics and Intelligent Systems (ICC-ROBINS), pp.607-612, Coimbatore, India,2024.

[10] K. Bhangale et al., "Machine Learning-based Heart Disease Prediction Using ECG Image," Proc. of the 2024 5th Int. Conf. for Emerging Technology (INCET), pp. 1-9, Belgaum, India, 2024.

[11] I. Farady, V. Patel, C. -C. Kuo and C. -Y. Lin, "ECG Anomaly Detection with LSTM-Autoencoder for Heartbeat Analysis," Proc. of the 2024 IEEE Int. Conf. on Consumer Electronics (ICCE), pp. 1-5, Las Vegas, NV, USA, 2024.

[12] M. L. Meghana et al., "Cardiovascular Disease Detection in ECG Images Using CNN-BiLSTM Model," Proc. of the 2024 IEEE Int. Conf. on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS), pp. 1-6, Bangalore, India, 2024.

[13] P. G. Aublin, J. Felblinger and J. Oster, "A Generalizable Heartbeat Classifier Leveraging Self-supervised Learning for ECG Analysis during Magnetic Resonance Imaging," IEEE Journal of Biomedical and Health Informatics, vol. 28, no. 9, pp. 5147-5155, Sept. 2024

[14] R. Krishna Priya, L. Alias and F. S. S. Al Salehiya, "Cardiac Health Assessment through Advanced Computational Models for ECG Image Analysis," Proc. of the 2024 10th Int. Conf. on Communication and Signal Processing (ICCSP), Melmaruvathur, India, 2024.

[15] Z. M. Tun and M. A. Khine, "Cardiac Diagnosis based ECG Images Classification System Using Convolution Neural Network," Proc. of the 2023 IEEE Conf. on Computer Applications (ICCA), pp. 387-392, Yangon, Myanmar, 2023.

[16] S. Krishnakumar et al., "Detection of Arrhythmia and Congestive Heart Failure Through Classification of ECG Signals Using Deep Learning Neural Network," Proc. of the 2021 Int. Conf. on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), Coimbatore, India, 2021.

[17] M. Moutaib, M. Fattah, Y. Farhaoui and B. Aghoutane, "Machine Learning in Fetal Health: Improving ECG Analysis with Random Forest," Proc. of the 2024 Int. Conf. on Circuit, Systems and Communication (ICCSC), pp. 1-5, Fes, Morocco, 2024.

[18] H. R. Esther T. et al., "Identification of Cardiovascular Disease Using ECG Images Based on Deep Learning Procedure," Proc. of the 2023 Int. Conf. on Advances in Computing, Communication and Applied Informatics (ACCAI), pp. 1-6, Chennai, India, 2023.

[19] J. E. Madias, "The 13th Multiuse ECG Lead: Shouldn't We Use It More Often, and on the Same Hard Copy or Computer Screen, As the Other 12 Leads?," Journal of Electrocardiology, vol. 37, no. 4, pp. 285-287, 2004.

[20] A. H. Khan and M. Hussain, "ECG Images Dataset of Cardiac Patients," Mendeley Data, V2, DOI: 10.17632/gwbz3fsgp8.2, 2021.
[21] R. C. Gonzalez and R. E. Woods, Digital Image Processing, 4th Edn., ISBN 978-0-13-335672-4, Pearson Education, 2018.
[22] C. Y. Saw and Y. C. Wong, "Neuromorphic Computing Based on Stochastic Spiking Reservoir for Heartbeat Classification," Jordanian Journal of Computers and Information Technology (JJCIT), vol. 8, no. 2, pp. 182-193, 2022.

ملخص البحث:

تُعدُّ أمراض القلب في طليعة أسباب الوفيات على مستوى العالم، وتتطّلب تشخيصاً دقيقاً ومبكّراً. نقدّم في هذه الدّراسة نظاماً لتصنيف أمراض القلب للاستفادة منه في تشخيص تلك الأمراض بالاعتماد على صُور تخطيط القلب وتقنيات التّعلُّم الآلي. تتألف الشبكة المقترحة من ثلاث خطواتٍ، تشمل: المعالجة الأوّلية، واستخلاص السِّمات، والتّعلُّم الجماعي.

في البداية تخضع صورة تخطيط القلب إلى معالجةٍ أوّليةٍ شاملة. يلي ذلك استخلاص السِّمات من الصّورة ومن ثم تقليل عددها باستخدام تحليل المكوِّنات الرئيسية (PCA) للحفاظ على السِّمات المميِّزة والمعلومات المهمّة المتضمَّنة في الصّورة. وفي نهاية المطاف، يتمّ استخدام نظامٍ مجمَّعٍ من عددٍ من نماذج التّعلُّم الآلي لتحسين أداء إطار العمل المقترح استناداً إلى طريقتي الانتخابٍ والتّرزيم.

وقد جرى تقييم النِّظام المقترح على مجموعة بياناتٍ عامّةٍ تشتملُ على صُور لتخطيط القلب، تمّ تصنيفها إلى أربع فئاتٍ: صُور طبيعية، وصُور غير طبيعية، وصُور تشير إلى نوبة قلبية، وصُور تشير إلى تاريخٍ من الإصابة بالنّوبات القلبية. وقد حقّق النّموذج المقترح أعلى دِقّة تصنيف تجاوزت 98% متفوِّقاً على النّماذج المماثلة الواردة في أدبيات الموضوع.

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 12, No. 01, March 2026.

GEA-COPE: AN EFFECTIVE MODEL FOR CROSS-DOMAIN GRAPH PRE-TRAINING

Yiming Zhao and Yongqing Wu

(Received: 10-Sep.-2025, Revised: 21-Nov.-2025, 10-Dec.-2025 and 13-Jan.-2026, Accepted: 14-Jan.-2026)

ABSTRACT

This paper addresses the negative transfer problem in cross-domain graph pre-training under few-shot learning scenarios, it proposes a multi-component pre-training framework called Graph External Attention-enhanced Coordinators for Pre-training (GEA-CoPe). This framework integrates multi-head external attention with a graph coordinator. Tackling the structural and semantic discrepancies between cross-domain graphs is crucial for mitigating negative transfer; however, conventional methods often lack adaptability to complex, dynamic inter-domain variations and explicit constraints for intermediate feature-distribution consistency. The proposed framework leverages an external attention-based coordinator to mediate between different graph datasets, dynamically generating cross-graph semantic-alignment strategies to alleviate negative transfer induced by structural heterogeneity. It employs a dual-feature normalization strategy that incorporates a cross-layer distribution alignment loss on top of intra-layer node-similarity constraints, effectively suppressing feature drift. Furthermore, Kolmogorov-Arnold Networks (KANs) are introduced, whose parameter-adaptive activation functions better capture non-linear topological dependencies and enhance model interpretability. Experiments on ten real-world graph datasets demonstrate that GEA-CoPe exhibits superior cross-domain generalization capability and significantly improves performance in few-shot node classification tasks, with an average improvement of about 13.3% *compared to other methods. The model can more accurately focus on critical graph structures, providing a theoretical foundation and practical paradigms for deploying graph neural networks in complex scenarios.*

KEYWORDS

Graph neural networks, Graph pre-training, Transfer learning, External attention.

1. INTRODUCTION

In recent years, in the fields of natural-language processing and computer vision, foundation models based on the Transformer architecture have acquired powerful general representation capabilities through pre-training on massive unlabeled data [1]. Subsequently, they can quickly adapt to various downstream tasks with minimal annotated data *via* fine-tuning, establishing a new "pre-training + fine-tuning" paradigm [2]. The success of this paradigm reveals the great potential of learning universal knowledge from large-scale data and transferring it to specific tasks. Inspired by this, the graph-learning community has also embarked on exploring the construction of "graph-foundation models," with cross-domain graph pre-training as their core component [3]. Cross-domain graph learning aims to train a universal graph encoder by integrating graph data from multiple sources with diverse structures and features, enabling it to learn transferable graph structural patterns and semantic knowledge transcending individual domains [4].

However, achieving this vision faces severe challenges. Real-world graph data exhibits extremely high heterogeneity. First, structurally, graphs from different domains may possess entirely distinct topological properties. For example, citation networks are typically homophilic [5], where connected nodes tend to belong to similar categories, whereas molecular networks or fraud-detection networks are often heterophilic, with connected nodes likely belonging to different categories. Second, at the feature level, node feature dimensions, physical meanings and distributions can vary significantly across different graphs [6]. This dual discrepancy in both structure and features makes effective knowledge transfer across different graph domains exceptionally difficult. Models are highly prone to learning knowledge on the source domain that cannot be applied to the target domain or even resulting in negative transfer [7]-[8]. Therefore, conducting research on cross-domain graph pre-training, exploring how to overcome the heterogeneity of graph data and building graph representation models capable of capturing universal patterns across domains, not only holds significant theoretical value, but is also an urgent

---

Y.-M. Zhao and Y.-Q. Wu (Corresponding Author) are with School of Software, Liaoning Technical University, Liaoning, China. Emails: 3130637271@qq.com and yqwuyywu@163.com

requirement for advancing graph intelligence technologies toward real-world applications.

Despite significant advances in cross-domain graph learning, existing methods still exhibit limitations when dealing with complex real-world graph data. Structure-oriented approaches [9]-[10] focus on mining commonalities in graph topology to achieve transferability through contrastive learning or structure generation. However, they often overlook the rich semantic information carried by node and edge features. When both the structure and the feature semantics differ significantly between the source and target domains, relying solely on structural similarity can lead to severe negative transfer. Feature-oriented methods [11]-[12] aim to align the feature spaces of different graph domains. Yet, they typically require consistent feature dimensions or depend on textual descriptions, which greatly restricts their applicability. For graph data with different feature dimensions or lacking explicit semantic annotations, feature alignment becomes particularly challenging. Hybrid approaches [13]-[15] often combine structure and feature information in a simple, sequential manner, failing to achieve deep and organic integration of both aspects. Furthermore, they struggle to effectively model global semantic relationships across graphs during training and are susceptible to feature-distribution shifts in deep networks, resulting in inefficient knowledge transfer and unstable model performance.

To address the aforementioned challenges, this paper proposes GEA-CoPe- an effective multi-component pre-training framework designed to alleviate negative transfer and feature drift in cross-domain graph learning. The framework demonstrates exceptional cross-domain generalization capability and significantly improves performance in few-shot node-classification tasks, achieving an average performance gain of approximately 13.3% compared to existing methods. It can be directly applied to cross-domain few-shot learning scenarios, such as transferring knowledge from a well-annotated citation network to classify nodes in a new social network or adapting a model from one e-commerce platform to another for user-interest recognition. Moreover, the framework serves as a robust foundational model for various downstream graph analytical tasks, particularly in target domains with limited supervisory signals. The main contributions of this work can be summarized as follows:

- An effective cross-domain graph pre-training framework is proposed. By leveraging a dynamic coordinator mechanism based on graph external attention, the model can implicitly learn deep semantic relationships across different graph domains. Through the dynamic interaction between coordinator nodes and external memory, it adaptively generates cross-graph semantic-alignment strategies, thereby effectively bridging domain gaps while preserving unique structural information of each graph, fundamentally mitigating negative transfer.
- A dual contrastive normalization module is designed to address feature drift in deep graph networks. It constrains feature smoothness among nodes within the same layer and ensures feature consistency during propagation through cross-layer distribution-alignment loss, enhancing the domain robustness and stability of pre-trained representations.
- In the downstream task-adaptation phase, Kolmogorov-Arnold Networks are introduced to replace traditional classification heads in cross-domain graph learning. With their superior non-linear fitting capability and higher parameter efficiency, KANs can better capture complex graph patterns and feature interactions, further improving the model's adaptability and generalization performance on target domains.
- Experiments were conducted on 10 datasets and the results proved the model’s superiority.

The remainder of this paper is structured as follows. Section 2 reviews related work. Section 3 elaborates on the proposed GEA-CoPe framework, which is structured into the pre-training phase and the transfer-learning phase. Section 4 describes the experimental setup and evaluation metrics, followed by a detailed presentation of the results. Finally, Section 5 concludes the paper and discusses its limitations along with potential directions for future research.

2. RELATED WORK

2.1 Graph Pre-training

Graph pre-training has emerged as a promising paradigm in graph machine learning. Its core idea is to leverage self-supervised learning on large-scale unlabeled graph data to capture universal structural and attribute patterns, thereby providing a well-initialized model with rich knowledge and strong generalization capability for downstream tasks. Typical pre-training strategies include node-level tasks, such as masked attribute reconstruction and context prediction, as well as graph-level objectives, like graph structure contrastive learning and property prediction. These tasks are designed to enable the model to deeply comprehend the complex dependencies among graph elements. Through such pre-training, the model can learn inherent and transferable domain knowledge, which significantly reduces dependence on labeled data in downstream tasks and effectively enhances generalization performance, convergence speed and final task performance. Current research primarily focuses on the following directions:

Contrastive Learning-based Graph Pre-training. Maximizes mutual information (MI) between graph structures or node sub-graphs to enhance the model's understanding of local and global feature correlations. For example, GraphCL [16] employs graph-augmentation strategies to generate multi-view contrastive samples, while SimGRACE [17] constructs positive-negative sample pairs *via* parameter perturbation to optimize node-level contrastive loss. These methods exhibit strong generalizability in molecular-property prediction and social-network analysis, but remain limited in modeling topological invariance for heterophilic graphs.

Generative Graph Pre-training. Forces models to learn the distribution patterns of graph data by reconstructing masked node attributes, edge connections or sub-graph structures. Representative methods include GPT-GNN [18], which adopts an auto-regressive approach to generate nodes and edges and GraphMAE [19], which introduces a masked auto-encoder to reconstruct node features. These methods excel in protein-interaction prediction but show low efficiency in reconstructing complex high-order relationships.

Cross-domain Universal Graph Pre-training Frameworks. For unified representation learning on multi-source heterogeneous graphs, recent studies proposed hierarchical contrastive pre-training [20], which separate domain-specific and shared features to enhance transferability in cross-domain tasks, like biomedicine and recommendation systems. However, challenges remain in integrating knowledge from large-scale heterogeneous graphs and adapting to temporal evolution in dynamic graphs.

2.2 Graph Transfer Learning

Graph transfer learning aims to transfer structural knowledge and semantic patterns learned from a source-graph domain to a target-graph domain to mitigate performance degradation caused by target-domain data scarcity or domain shifts. Its core challenge lies in aligning cross-domain topological heterogeneity and extracting domain-invariant representations. Recent research directions include:

Domain Adaptation-based Graph Transfer. Reduces structural discrepancies between source and target domains *via* adversarial training or distribution alignment. For instance, [21] introduced a graph convolutional adversarial framework that jointly aligns node features and topological structures by minimizing domain divergence through Wasserstein distance constraints, while [22] formalized Fused Gromov-Wasserstein distance for structured graph alignment, providing theoretical foundations for minimizing inter-domain Wasserstein distances. These methods perform robustly in cross-social network user-behavior prediction, but struggle to adapt to temporal dynamics in dynamic graphs.

Heterogeneous Graph Representation Transfer. Meta-path-aware transfer frameworks address node/edge type heterogeneity. The Heterogeneous Graph Transformer [23] dynamically adjusts relation-specific attention weights through meta-relation aware mechanisms, while [24] employs reinforcement learning for automated meta-path discovery across domains. These methods demonstrate effectiveness in cross-platform recommendation tasks without relying on predefined-schema constraints, as validated in Amazon eBay product alignment experiments.

Dynamic Graph Temporal Transfer. Recent advances handle structural evolution through temporal modeling. [25] decouples graph-convolution parameters into temporal trajectories using RNNs, capturing both topological persistence and variation patterns. [26] implements continuous-time graph representation learning *via* temporal point processes, effectively addressing domain shifts in financial-transaction networks with adaptive computation.

Unsupervised Cross-graph Transfer. Domain-invariant feature-learning methods achieve progress through novel objectives. DANE [27] disentangles domain-specific variations *via* adversarial alignment of graph embeddings, while Graph Optimal Transport [28] maximizes feature correspondence through

Wasserstein-distance minimization. These approaches show superior performance in cross-organism protein network analysis with explicit geometric-alignment constraints.

3. METHOD

Our pre-training dataset consists of M graphs, represented as $\mathcal{G}^{(i)} = \left(\mathcal{V}^{(i)}, \mathcal{E}^{(i)}\right)$, where $i \in \{1,2,\dots,M\}$, respectively. $\mathcal{V}^{(i)} = \left\{v_1^{(i)}, v_2^{(i)}, \dots, v_{|\mathcal{V}^{(i)}|}^{(i)}\right\}$ and $\mathcal{E}^{(i)} = \mathcal{V}^{(i)} \times \mathcal{V}^{(i)}$ represent the node sets and edge sets, respectively. Each $\mathcal{G}^{(i)}$ graph is associated with a feature matrix $X^{(i)} \in \mathbb{R}^{|\mathcal{V}^{(i)}|\times d_i}, E^{(i)} \in \mathbb{R}^{|\mathcal{E}^{(i)}|\times d_i}$ and an adjacency matrix $A^{(i)} \in \mathbb{R}^{|\mathcal{V}^{(i)}|\times|\mathcal{V}^{(i)}|}$. The main goal is to train a graph neural network(GNN) $h(\cdot)$ with learnable parameter Θ that captures domain-agnostic knowledge for adaptation to downstream applications. The downstream dataset is represented as $\mathcal{G}^{(t)} = \left(V^{(t)}, \mathcal{E}^{(t)}\right)$ with the feature matrix $X^{(t)}$ and adjacency matrix $A^{(t)}$.

3.1 Overview of Our Framework

In this sub-section, the proposed GEA-CoPe model is described in detail, which consists of two phases. In the first phase, pre-training is conducted on multiple cross-domain graph datasets and the established pre-training frameworks GraphCL [16] and SimGRACE [17] are used to guide the entire process. The second phase implements transfer learning to adapt the pre-trained knowledge to downstream tasks for addressing diverse applications. Several novel techniques are introduced to address the aforementioned issues and challenges, with the overall framework illustrated in Figure 1.

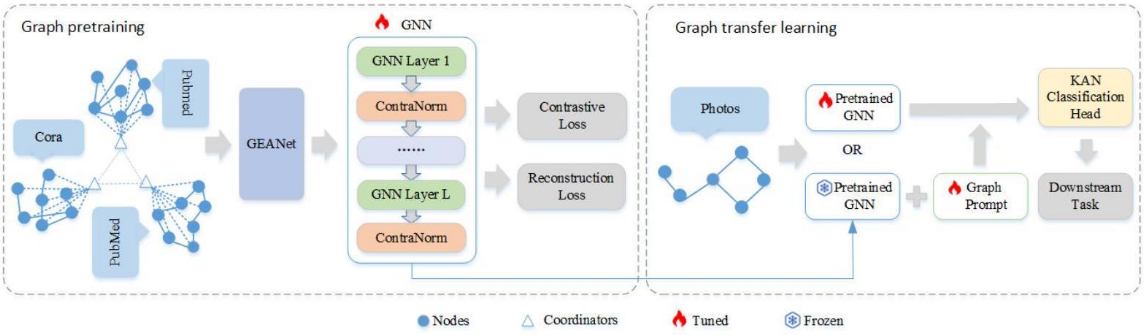


Figure 1. Overall framework of GEA-CoPe model. The first half is the graph pre-training stage and the second half is the graph transfer-learning stage.

3.2 Aligning Graphs by Coordinators

To address the heterogeneous feature representations and topological disparities across graph data, this paper uses an alignment framework. This architecture comprises two core stages: First, feature-space standardization transforms heterogeneous node features into a unified dimensional space through linear projections. Subsequently, a dynamic coordinator mechanism introduces learnable virtual nodes to establish cross-graph semantic correlations, enabling dual-level adaptive alignment of structural patterns and semantic relationships. This phased methodology systematically resolves both shallow feature-distribution discrepancies and deep pattern-expression variations in cross-domain graph datasets.

3.2.1 Data Pre-processing

During the data pre-processing stage, a series of crucial steps is performed to ensure the effectiveness of cross-domain graph pre-training. First, raw graph data is loaded from ten standard graph datasets and uniformly converted into a data object format, achieving standardized integration of multi-source data. Subsequently, data-cleaning operations are executed to remove pre-defined data-split masks (including training, validation and test-set identifiers) from the datasets. This step effectively prevents potential data-leakage risks and provides a clean data foundation for subsequently constructing a unified cross-dataset pre-training paradigm. Next, unified dimensionality processing is applied to each graph. When the original feature dimensionality (e.g. 1433 dimensions for Cora) is higher than the target dimensionality (100 dimensions), Singular Value Decomposition (SVD) is employed for dimensionality reduction, preserving over 95% of the variance. When the original feature dimensionality is lower than the target, zero-padding is performed to ensure consistent node feature dimensions across all graphs. Following this, multiple independent graphs are merged into a unified large graph. For each original graph, a learnable coordinator node is added, generating learnable features. Edge-connection strategies include static full connection (increasing the edge count by 57.1%) and dynamic similarity-based connection. When dynamically adding edges based on node feature cosine similarity, only edges with a similarity above a threshold (default 0.1) are retained to control sparsity (resulting in edge-count increases ranging from 7.6% to 56.8%). Finally, sub-graphs are sampled from the large graph *via* a random walk algorithm, with a walk length of 30 and starting nodes comprising 10% of the total, covering $70\% - 85\%$ of all nodes. Sub-graphs with fewer than 5 nodes are filtered out to ensure sample quality, forming the final collection of sub-graphs for subsequent learning.

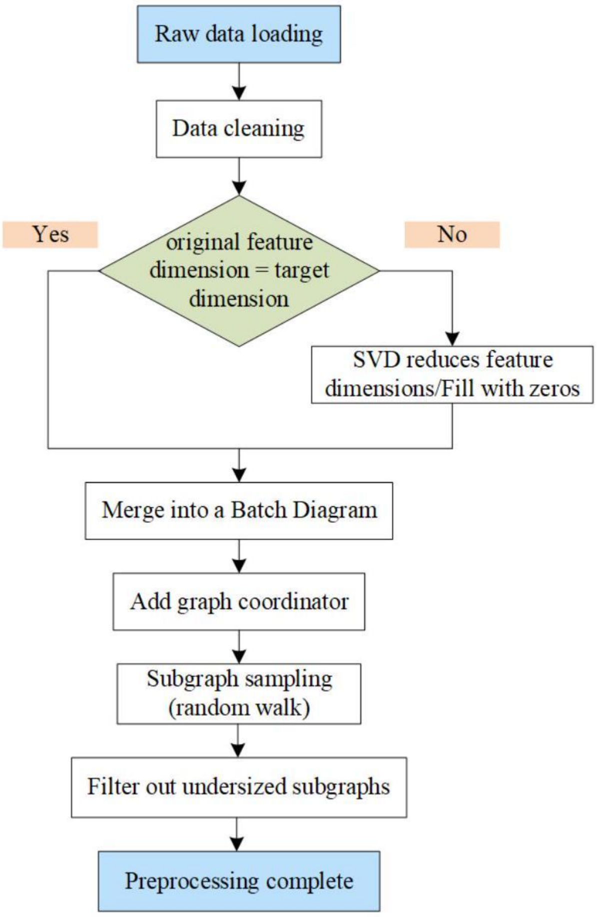

Raw data loading
Data cleaning
Yes
original feature dimension = target dimension
No
SVD reduces feature dimensions/Fill with zeros
Merge into a Batch Diagram
Add graph coordinator
Subgraph sampling (random walk)
Filter out undersized subgraphs
Preprocessing complete

Figure 2. Data pre-processing flowchart.

The data pre-processing flowchart is illustrated in Figure 2. These pre-processing steps collectively form a systematic pipeline that transforms multiple graph datasets into a structurally unified and feature-aligned collection of sub-graphs. This pipeline provides a solid data foundation for the subsequent cross-domain graph pre-training model, ensuring the model's robustness and generalization capability when processing graph data.

3.2.2 Feature Projection

In the first stage of the method; namely, the pre-training phase, the initial step involves processing the data to align the feature dimensions across different domains, as shown in Figure 3. This is achieved through a projection module, with the specific implementation as follows:

$$\tilde{X}^{(i)} = \mathrm{Proj}\left(X^{(i)}\right) \in \mathbb{R}^{|\mathcal{V}^{(i)}| \times dp}, \tag{1}$$

where $\mathrm{Proj}()$ denotes the projection operation and d_p denotes the pre-defined projected dimension. In this paper, the widely-addressed singular value decomposition (SVD) is employed for the projection operation. However, to address the feature-alignment problem, merely applying feature projection to the data is insufficient; additional calibration processes are required to further calibrate the data.

“GEA-CoPe: An Effective Model for Cross-domain Graph Pre-training”, Y.-M. Zhao and Y.-Q. Wu.

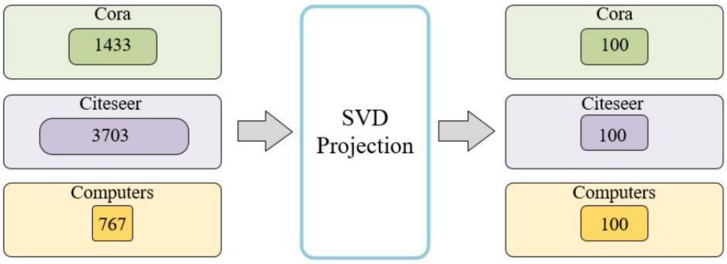


Figure 3. SVD feature projection.

3.2.3 Graph Coordinators

Following the feature-projection stage described above, a "coordinator" - virtual node is introduced, which is designed to bridge graphs from different domains and enhance feature and structural alignment.

Coordinator-Graph Connection. Considering that datasets originate from distinct domains where each graph exhibits unique structural properties and information flows, in order to preserve the intrinsic structural characteristics of individual graphs while enabling their participation in cross-graph information exchange, a dedicated coordinator is established for each dataset. Rather than being isolated from the graph data, each coordinator is fully connected to all nodes within its associated graph, forming a new sub-graph that becomes an integral part of the original graph. This design creates direct and efficient communication pathways between the coordinator and nodes, allowing the coordinator to effectively gather node-level information and facilitate coordinated interactions.

Coordinator-Coordinator Connection. Since our objective focuses on enabling cross-domain knowledge sharing rather than enhancing individual graph representations, inter-coordinator connections are established to serve as bridges for inter-graph communication. Specifically, edges are introduced between coordinators originally assigned to different graph datasets, thereby constructing inter-connected channels for global information exchange. This eliminates data isolation and creates a unified platform for comprehensive knowledge sharing across all domains. Through these operations, a joint adjacency matrix is constructed, including the original graph adjacency matrix and the newly added coordinator connection. The formula is:

$$\tilde{A} = \begin{bmatrix} A_{\text{diag}} & R_A^T \\ R_A & R_R \end{bmatrix} \tag{2}$$

where $A_{\text{diag}} = \text{Diag}\left(A^{(1)}, A^{(2)}, \dots, A^{(M)}\right), R_A = \text{Stack}\left(R_A^{(1)}, R_A^{(2)}, \cdots, R_A^{(M)}\right), R_R = 1^{M\times M}$. Diag means concatenating matrices diagonally and Stack means stacking row-vectors into a matrix. $R_A^{(i)} \in \mathbb{R}^N$, the j th value of $R_A^{(i)}$:

$$N = \sum_{k}^{M} \left|v^{(k)}\right|, \tag{3}$$

$$R_A^{(i)}(j) = \begin{cases} 1, \sum_{1}^{i} \left|v^{(k)}\right| \leq j < \sum_{1}^{i+1} \left|v^{(k)}\right|, \\ 0, \text{otherwise} . \end{cases} \tag{4}$$

The coordinator representation serves as a learnable parameter that can be trained jointly with GNNs. Through an end-to-end collaborative training design, adaptive units dynamically calibrate their topological connection weights and feature-aggregation patterns according to the underlying data distribution. This collaborative training mechanism ensures that the coordinator continuously self-improves as an information bridge, enabling more effective transmission of cross-domain graph knowledge.

Generate Graph Batches for Efficient Training. By strategically leveraging coordination mechanisms to bridge disparate graph structures, this framework implements cross-graph node sampling during training iterations. Such synergistic processing enhances pre-training through batch-level knowledge amalgamation while promoting cross-dataset feature alignment. The co-optimization paradigm compels the model to distill topological regularities transcending individual graph boundaries, thereby deriving unified latent representations that comprehensively synthesize graph information from diverse domains. The cross-fusion of graph characteristics during parameter updates establishes an inductive bias favoring the extraction of fundamental relational patterns while simultaneously facilitating the advancement of cross-domain graph-learning frameworks through coordinated structural integration.

3.3 Graph External Attention

The self-attention mechanism assumes that the input graph is fully connected. Initially, each element of the input sequence is transformed into vector representations *via* an embedding layer. Each input vector is then linearly projected to generate three vectors: the query (Q), which explores correlations with other positions; the key (K), which is matched by queries from other positions; and the value (V), which stores the actual information to be aggregated. These operations are formally expressed as:

$$\mathrm{A}_{\mathrm{Self}} = \mathrm{softmax}\left(\frac{\mathrm{QK}^{T}}{\sqrt{d_{\mathrm{out}}}}\right) \in \mathbb{R}^{|v^{(i)}|\times|v^{(i)}|} \tag{5}$$

$$\mathrm{Self} - \mathrm{Attn}(\mathrm{X}) = \mathrm{A}_{\mathrm{Self}}\,\mathrm{V} \in \mathbb{R}^{|v^{(i)}|\times d_{\mathrm{out}}} \tag{6}$$

where, W_Q, W_K, W_V represent trainable parameters and d_{out} denotes the dimension of Q. Subsequently, attention scores are computed to perform weighted aggregation, where positions with higher relevance are assigned greater weights. However, conventional self-attention mechanisms predominantly focus on node features within a single graph and capture only superficial associations between nodes, which limits their functional capabilities.

Inspired by [29], a graph external attention network is introduced, which not only attends to node features within individual graphs, but also incorporates external units. By computing attention between these external units and the node features of the input graph, the proposed method enhances graph representation learning. This approach achieves:

$$\mathrm{A}_{GE} = \mathrm{norm}(\mathrm{XU}^{T}) \in \mathbb{R}^{|v^{(i)}|\times S}, \tag{7}$$

$$\mathrm{GE} - \mathrm{Attn}(\mathrm{X}) = \mathrm{A}_{GE}\mathrm{U} \in \mathbb{R}^{|v^{(i)}|\times d_i}, \tag{8}$$

where, $\mathrm{U} \in \mathbb{R}^{S\times d_i}$ as external units, is designed as learnable parameters containing S nodes, with the information being shared across all input graph data. A_{GE} denotes the similarity between the input-graph nodes and the external units. Subsequently, normalization operations [47] are applied to A_{GE}, specifically performing row-wise and column-wise normalization, respectively. To elaborate, the normalization is implemented by:

$$\tilde{\alpha}_{i,j} = (\mathbf{XU}^{T})_{i,j}, \tag{9}$$

$$\hat{\alpha}_{i,j} = \frac{\exp(\tilde{\alpha}_{i,j})}{\sum_{k=0}^{n} \exp(\tilde{\alpha}_{k,j})}, \tag{10}$$

$$\alpha_{i,j} = \frac{\hat{\alpha}_{ij}}{\sum_{k=0}^{S} \hat{\alpha}_{i,k}}. \tag{11}$$

In specific implementations, to achieve enhanced performance, two external modules are used to store keys and values respectively. Furthermore, a separate external module is utilized to process edge features within the input graph, while node-edge connectivity information is incorporated into a shared module.

$$\mathrm{X}_{out} = \mathrm{norm}\left(\mathrm{XU}_{s}\mathrm{U}_{nk}^{T}\right)\mathrm{U}_{nv}, \tag{12}$$

$$\mathrm{E}_{out} = \mathrm{norm}\left(\mathrm{EU}_{s}\mathrm{U}_{ek}^{T}\right)\mathrm{U}_{ev}, \tag{13}$$

where $\mathrm{U}_s \in \mathbb{R}^{|v^{(i)}|\times|v^{(i)}|}$ represents shared units; $\mathrm{U}_{nk}, \mathrm{U}_{nv} \in \mathbb{R}^{S\times d_i}$ is the external unit for storage nodes, while $\mathrm{U}_{ek}, \mathrm{U}_{ev} \in \mathbb{R}^{S\times d_i}$ is the external unit for storage edges.

The multi-head self-attention mechanism serves as the core component of Transformer models, with its fundamental principle being the processing of input sequences through multiple parallel self-attention modules followed by result integration to enhance the model's expressive power. For instance, both the node-node relationships within a graph and the node-external unit relationships exhibit complex diversity. Therefore, the processing analogous to the multi-head self-attention mechanism is adopted:

$$h_i = \mathrm{GE} - \mathrm{Attn}(\mathrm{X}_i, \mathrm{U}_{nk}, \mathrm{U}_{nv}) \tag{14}$$

$$\mathrm{X}_{out} = \mathrm{MultiHeadGEA}\ (\mathrm{X}, \mathrm{U}_{nk}, \mathrm{U}_{nv}) = \mathrm{Concat}(h_1, \ldots, h_H)\mathrm{W}_o \tag{15}$$

where $\mathbf{U}_{nk}, \mathbf{U}_{nv} \in \mathbb{R}^{S \times d_i}$ denotes the memory unit shared by all heads. h_i represents the i-th head, H represents the total number of heads and W_o is a linear transformation matrix Finally, a skip connection is applied to the output.

3.4 Pre-training on Multi-domain Graphs

This paper proposes a universal cross-domain graph pre-training framework compatible with various pre-training methods, which generates more expressive embeddings at both node and graph levels. Existing works predominantly focus on paradigms utilizing homogeneous data domains for pre-training [30]. GraphCL [16] systematically constructs positive sample pairs through structured graph data-augmentation strategies, explicitly enhancing data diversity to guide models in learning invariant features while maximizing mutual information between augmented and original samples through contrastive loss. SimGRACE [17] directly generates positive sample pairs by applying subtle perturbations to GNN encoder parameters. Since parameter perturbations preserve the topological connectivity of original graph structures, they comprehensively retain global graph attributes. Based on these considerations, GraphCL and SimGRACE were selected as our pre-training methods.

During pre-training, significant mean and variance discrepancies in node features across different GNN layers lead to gradual dilution of shallow semantic information in deeper layers. Traditional methods exacerbate feature-distribution oscillation in few-shot scenarios due to biased mini-batch statistical estimations. To address these issues, ContraNorm [31] is introduced, which is a systematic solution. Conventional normalization techniques solely focus on single-layer feature distributions, whereas our dual contrastive-normalization method incorporates dual optimization objectives: intra-layer feature smoothness and cross-layer distribution consistency. By synchronously implementing feature-space compactness and inter-layer distribution alignment after each GNN layer, expressed as:

$$H_t = \text{ LayerNorm } \left(H_b - \frac{s}{\tau} \times \text{softmax}\left(H_b H_b^{\top}\right) H_b\right), \tag{16}$$

Where H_b and H_t represent the feature matrices before and after the update, respectively, s denotes the step size of gradient descent and τ is the temperature.

To ensure the integrity of graph structural information, this framework introduces an auxiliary feature-reconstruction loss. The loss is measured through Mean Squared Error (MSE), which quantifies the preservation of node-feature information by computing the MSE between linearly transformed raw node-feature vectors and reconstructed feature vectors. Specifically, the framework employs MLP to decode low-dimensional node embeddings, generating reconstructed features aligned with the original feature space. This mechanism aims to achieve dual objectives: at the single-graph level, it preserves crucial node characteristics during dimensionality reduction; at the multi-graph alignment level, it enhances compatibility among different graph-embedding spaces through feature-fidelity constraints, thereby mitigating information redundancy caused by feature-distribution discrepancies in cross-graph tasks. Taking GraphCL as an example, the pre-training objective is formulated as:

$$\mathcal{L} = -\log \frac{\exp\left(\sin\left(h\left(\text{PS}(\tilde{X}, \tilde{A}, a_i)\right)\right), h\left(\text{PS}(\tilde{X}, \tilde{A}, a_j)\right) / \tau\right)}{\sum\ \exp\left(\sin\left(h\left(\text{PS}(\tilde{X}, \tilde{A}, a_i)\right)\right), h\left(\text{NS}(\tilde{X}, \tilde{A}, a_j)\right) / \tau\right)} + \|\tilde{X} - \hat{X}\|_2, \tag{17}$$

where $\tilde{X}$ denotes the feature matrix formed by concatenating all pre-training datasets, $\tilde{A}$ represents the adjacency matrix connected *via* the coordinator, PS and NS correspond to Positive Sampling and Negative Sampling, respectively, sim indicates the similarity measurement, a_i and a_j are two distinct graph-augmentation methods, λ serves as the reconstruction loss coefficient governing the emphasis on the reconstruction task.

3.5 Applying Knowledge to Downstream Data

Our pre-training method GEA-CoPe demonstrates compatibility with diverse techniques through its task agnostic nature and task-space adaptability. During the transfer phase, the node classification is selected as the downstream task, where conventional approaches typically employ MLP as classification heads. This paper proposes replacing MLP with KAN, the core advantage of which lies in dynamically capturing complex non-linear relationships between node features through a kernel attention mechanism. KAN [32] explicitly models node similarity through this mechanism, proving particularly effective for heterophilic graphs. The incorporation of sparse attention mechanisms reduces computational overhead while maintaining suitability for large-scale graph data. During cross-domain transfer, attention weights adaptively adjust feature importance to mitigate inter-domain distribution discrepancies.

Building upon recent advancements in graph neural networks [33], a graph-level framework is constructed for downstream tasks. Since knowledge-transfer efficiency improves when pre-training tasks and downstream applications maintain topological-space alignment, both stages employ graph-level representations. Specifically, adjacency-matrix reconstruction techniques are implemented to lift node-level tasks to the graph space, as detailed in Algorithm 1.

Algorithm 1: GEA-CoPe

```
Input: Source graphs $\{\mathcal{G}^{(i)}\}_{i=1}^{M}$, target graph $\mathcal{G}^{(t)}$, GNN parameters $\Theta$, projection operation $\mathrm{Proj}(\cdot)$, pre-training objective $\mathcal{L}(\cdot)$, learning rate $\alpha$, transferring pipeline $\mathrm{Trans}(\cdot)$
Output: The optimal model on the target graph $g_t(\cdot)$
for $i \leftarrow 0$ to $M$ do
    $\tilde{X}^{(i)} = \mathrm{Proj}\left(X^{(i)}\right)$
end for
$\tilde{X} = \mathrm{Cat}\left(\tilde{X}^{(1)}, \tilde{X}^{(2)}, \ldots, \tilde{X}^{(M)}\right)$
$\tilde{A} = \begin{bmatrix} A_{\mathrm{diag}} & R_A^T \\ R_A & R_R \end{bmatrix}$
while not converge do
    $\Theta \leftarrow \Theta - \alpha \nabla_{\Theta} \mathcal{L}\left(\tilde{X}, \tilde{A}, \Theta\right)$
end while
$g_t(\cdot) = \mathrm{Trans}\left(\mathcal{G}^{(t)}, \Theta\right)$
return $g_t(\cdot)$
```

3.6 Complexity Analysis

The feature complexity of the coordinator is $O(Md_{\mathrm{A}})$, exhibiting a linear relationship with the number of pre-training datasets. In practical scenarios, situations with a large number of pre-training datasets are rare. Assuming that the employed GNN comprises L layers with a maximum layer width of d and letting $N = \sum_{k=1}^{M} \left|\mathcal{V}^{(k)}\right|$ and $E = \sum_{k=1}^{M} \left|\mathcal{E}^{(k)}\right|$, the computational cost of GEANet scales linearly with the number of nodes and edges, with a complexity of $O(N + E)$. It is noteworthy that the time complexity of a typical graph model (e.g. Graph Convolutional Network, GCN) is $O(LNd^2 + LEd + Nd)$. After incorporating the coordinator, the time complexity becomes $(L(N + M)d^2 + L(E + N + M)d + (N + M)d + (N + E))$, with an additional time complexity of $O(LMd^2 + L(N + M)d + Md + (N + E))$. When $M \ll N$, the $O(N + E)$ term from GEANet is incorporated into the linear terms of the coordinator. The dominant term remains $O(LNd^2)$ from the GNN layers and the supplementary time cost exhibits an approximately linear relationship with the original number of nodes.

4. EXPERIMENTS

In this section, experiments are conducted on various graph datasets to evaluate the methods proposed in this paper and the baseline methods and analyze the experimental results. All experiments were conducted on a server equipped with a single NVIDIA GeForce RTX 3080 GPU (10 GB memory), an Intel Xeon Platinum 8352 V CPU (12 cores @ 2.10 GHz) and 48 GB of RAM. The software environment consisted of the Ubuntu 22.04 operating system, PyTorch 2.1.2 deep-learning framework, Python 3.10 programming language and CUDA 11.8 parallel-computing platform. During the training phase, a batch size of 100 was used, with training proceeding for 100 epochs and a total training time of approximately 1.5 hours.

4.1 Experimental Setup

4.1.1 Dataset

To evaluate the accuracy of the assessment, experiments were conducted on ten real-world benchmark datasets. These datasets include five homophilic datasets: Cora [34], Citeseer [34], Pubmed [35], Computers and Photos [36]-[37], as well as five heterophilic datasets: three sub-datasets from WebKB [38] (Cornell, Texas and Wisconsin) and two page networks extracted from Wikipedia [38] (Chameleon and Squirrel). Detailed information is presented in Table 1, where the values from [39] are used to measure the degrees of homophily and heterophily. As shown in the table, the first five datasets exhibit strong homophily, while the latter five demonstrate significant heterophily [39]-[40]. The varying degrees of homophily and heterophily reflect distinct semantic representations in graph structures.

Table 1. Statistics of datasets.

Homophilic Data	Cora	Citeseer	Pubmed	Computers	Photos
#Nodes	2,708	3,327	19,717	13,752	7,650
#Edges	10,556	9,104	88,648	491,722	238,162
#Features	1,433	3,703	500	767	745
#Labels	7	6	3	10	8
$h(G)$	0.810	0.736	0.802	0.777	0.827
Heterophilic Data	Wisconsin	Texas	Cornell	Chameleon	Squirrel
#Nodes	251	183	183	2,277	5,201
#Edges	515	325	298	62,792	396,846
#Features	1,703	1,703	1,703	2,325	2,089
#Labels	5	5	5	5	5
$h(G)$	0.196	0.108	0.305	0.231	0.222

4.1.2 Baselines

To evaluate the performance of GEA-CoPe, the framework is compared with the following baselines, which are broadly categorized into three groups and briefly summarized.

Supervised Methods: These approaches typically train GNN models on downstream tasks for direct inference. In this study, two widely-used GNN architectures are implemented: GCN [41] and FAGCN [42]. These models are selected as the backbone of our proposed GEA-CoPe method, because FAGCN is specifically tailored for both homophilic and heterophilic graphs [39], while GCN serves as a widely-used foundational GNN model that underpins FAGCN.

Isolated Pre-training with Fine-tuning: These methods leverage multiple cross-domain datasets as source datasets, which are combined in an isolated manner to pre-train GNN models in a self-supervised fashion (e.g. GraphCL [16] and SimGRACE [17]). Here, "isolated" indicates that the datasets are merged into a single batch object, resulting in an adjacency matrix composed of distinct blocks. Subsequently, the pre-trained model is fine-tuned for new downstream tasks.

Graph Coordinator for Pre-training (GCOPE) [45]: This methodology integrates disconnected source datasets into a unified large-scale graph through a coordination mechanism that establishes cross-dataset dependencies during pre-training. The resulting model is then transferred to downstream applications.

External Attention-Augmented Graph Coordinator for Pre-training (GEA-CoPe): Our proposed method employs an external attention-augmented learnable coordinator to act as a bridge for information interaction across diverse graph datasets. The pre-trained GNN model is then transferred to downstream tasks through fine-tuning or prompting, alleviating the negative-transfer problem [15].

4.1.3 Metrics and Implementations

Three universally adopted metrics were selected for evaluating node-classification tasks [39], [43]-[44]: classification accuracy (Acc), mean AUC-ROC value (AUC) and mean F1-score (F1). A 10-fold partition strategy was applied to divide ten real-world benchmark datasets, with nine serving as cross-domain source datasets for model pre-training and the remaining one designated as the target domain for transfer learning. To harmonize feature-distribution discrepancies across multiple cross-domain sources, SVD was employed for dimensionality reduction, compressing original features to 100 dimensions. Subsequently, an independent coordination module is assigned to each source dataset, with the default reconstruction-weight coefficient set to 0.2. For the external-attention module, the number of attention heads is set to 4.

In the pre-training phase, a contrastive learning framework is adopted. The number of graph neural network layers is set to 8 and the hidden dimension is set to 128. Standard dropout regularization is applied to prevent overfitting; a dropout rate of 0.2 is used to enhance model robustness to some extent while avoiding excessive loss of information flow [46]. All networks are optimized with the Adam optimizer, with a base learning rate uniformly set to 0.0001 to ensure stable learning of general representations. Weight decay is set to 0.00001, which prevents overfitting without unduly weakening the model's expressive power.

In the transfer-learning phase, node classification serves as the primary downstream task and the training sets are constructed following the C-way-K-shot few-shot learning paradigm described in reference [48]. The remaining data is randomly split into validation and test sets in a 1: 9 ratio. The split_ratio is set to 0.1, indicating that 10% of all nodes are randomly selected as starting nodes for random walks, with each random walk length set to 30. A split ratio of 0.1 better simulates the scarcity of labeled data in real-world scenarios, thereby more effectively evaluating the model's generalization ability.

Table 2. Hyper-parameter settings.

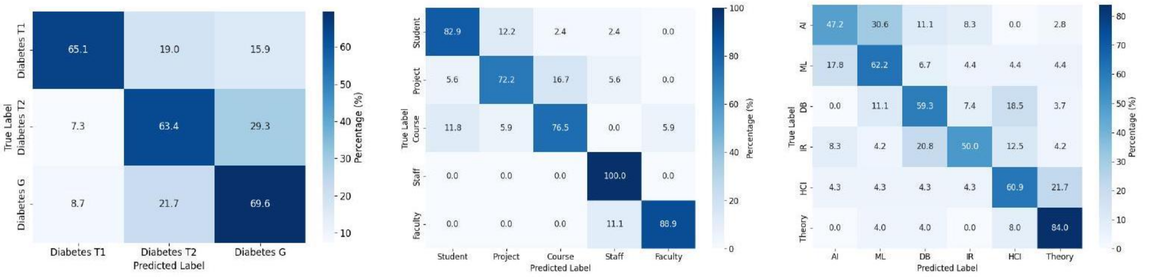
Hyper-parameter	Value
Node Feature Dimension	100
Reconstruction Loss Weight	0.2
Number of Attention Heads	4
Number of Convolutional Layers	8
Hidden Dimension	128
Dropout Rate	0.2
Optimizer(Learning Rate)	Adam(1e-4)
Optimizer(Weight Decay)	Adam(1e-5)
Random Walk Split Ratio	0.1
Random Walk Length	30

The hyper-parameter settings are listed in Table 2. To ensure robust performance across datasets and avoid performance degradation, the pre-training phase prioritizes tuning the learning rate and batch size to guarantee stable convergence, then gradually introduces reconstruction loss weight and dynamic edge pruning thresholds to enhance generalization. During fine-tuning, the learning rate is adjusted dynamically according to the sample size of the downstream task; in few-shot scenarios, the batch size is reduced and the number of training epochs is increased. The number of neural-network layers is adjusted based on the graph diameter and signs of overfitting are monitored to regulate the dropout rate. The Adam optimizer is employed throughout the experiments. Only 1-2 hyper-parameters are adjusted at a time, with evaluation *via* cross-validation. When transferring across datasets, adaptive adjustments are made according to differences in graph scale and feature distribution between the source and target domains.

4.2 Few-shot Performance Evaluation

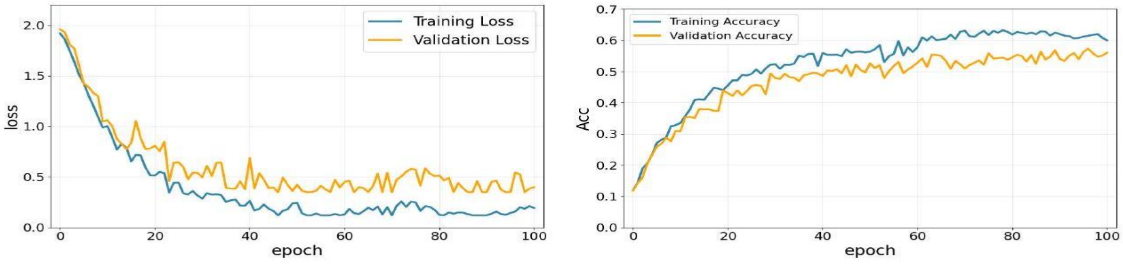
The GEA-CoPe was compared with three baseline groups on node-classification tasks under the C-way-1shot setting. Results on homophilic graph datasets are presented in Table 3, while those on heterophilic graphs are shown in Table 4. By analyzing the performance of supervised-learning methods, the effectiveness of pre-training GNN transfer is verified and the necessity of knowledge transfer is demonstrated. Undoubtedly, the core objective of pre-training lies in learning universal features or knowledge from large-scale data to provide foundational models for downstream tasks, thereby enhancing model performance, efficiency and generalization capabilities, particularly under few-shot conditions.

Based on our findings, the performance of supervised methods is notably inferior, with negative transfer being particularly prominent. The primary issue stems from the substantial divergence in data structures and distributions across datasets from different domains. During pre-training, samples contain information from only a single dataset and remain isolated; consequently, they fail to integrate comprehensive graph information. This consequently leads to compromised effectiveness in GNNs' learning of graph representations. It is observed that IP with fine-tuning often fails to achieve performance comparable to supervised methods, manifesting as the negative-transfer phenomenon. This is attributed to significant distribution shifts across different source domains. Under the IP strategy, each graph sample originates from one of nine distinct data distributions. As a result, graph neural networks struggle to reconcile these disparate distributions into a unified representation space, thereby limiting their ability to learn generalizable graph features. Although the GCOPE method with graph coordinator connects cross-domain graphs into a unified framework, enabling better representation learning across graphs, its lack of effective feature-enhancement modules for node and edge attributes constrains model expressiveness, resulting in unstable feature distributions and weak generalization. In contrast, our proposed GEA-CoPe method significantly outperforms these baselines. The incorporated multi-head self-attention mechanism enhances data representation by enabling simultaneous focus on diverse feature sub-spaces, distributing attention focus and mitigating single-attention bias. Through attention-driven feature enhancement and structured computational optimization, our method improves both accuracy and efficiency, upgrading the coordinator from a basic parameter-matching framework to an efficient universal processor suitable for complex graph-structured data. Therefore, during pre-training, our approach enables more effective integration of multi-dataset information and enhances graph representations for downstream applications.



(a) (b) (c)

Figure 4. Node-classification confusion matrix of GEA-CoPe (c-way-1-shot). (a)Confusion matrix of node classification on Cora. (b)Confusion matrix of node classification on Texas. (c)Confusion matrix of node classification on Citeser.



(a) (b)

Figure 5. Node-classification accuracy and loss of GEA-CoPe on PubMed (c-way-1-shot). (a)Accuracy curve. (b)Loss curve.

Additionally, to more intuitively demonstrate the framework's performance, partial confusion matrices are plotted. As shown in Figure 4, which displays the classification results on the Cora, Texas and Citeseer datasets from left to right, the distribution within the confusion matrices reveals that the framework demonstrates significant advantages in multi-class classification tasks, particularly exhibiting strong robustness when handling complex feature interactions and ambiguous class boundaries. In the diabetes-type classification task, the framework's high accuracy for Gestational Diabetes (69.6%) reflects its strong ability to identify categories with distinct feature differences. In the user-role classification task, the perfect identification of the Staff category (100%) indicates the framework's effectiveness in capturing the unique patterns of minority or distinctively featured classes, showcasing its adaptability to extremely distributed data. In the academic-domain classification task, the high accuracy for the Theory category (84.0%) confirms the framework's capability for hierarchical modeling of classes with clear semantic features. Overall, through multi-dimensional feature-decoupling and contextual-relationship modeling, the framework efficiently identifies well-separated categories while clearly exposing bottlenecks related to feature overlap and label ambiguity. The accuracy and loss variations of the framework in node classification on PubMed are shown in Figure 5.

Furthermore, to evaluate the framework's competitiveness in cross-domain graph learning, four state-of

Table 3. Transfer learning performance (mean±std Acc/AUC/F1) on homophilic datasets (C-way-1-shot). GCL and Sim respectively represent GraphCL and SimGRACE.

Training schemes	Methods	Cora			Citeseer			Pubmed			Computers			Photos		
		Acc	AUC	F1	Acc	AUC	F1	Acc	AUC	F1	Acc	AUC	F1	Acc	AUC	F1
Supervised	GCN	$0.3027_{\pm.06}$	$0.6436_{\pm.06}$	$0.2783_{\pm.07}$	$0.3760_{\pm.04}$	$0.7230_{\pm.03}$	$0.3280_{\pm.04}$	$0.3959_{\pm.01}$	$0.5443_{\pm.02}$	$0.3575_{\pm.08}$	$0.2537_{\pm.07}$	$0.6602_{\pm.01}$	$0.2289_{\pm.04}$	$0.4092_{\pm.04}$	$0.7817_{\pm.04}$	$0.3849_{\pm.07}$
	FAGCN	$0.3359_{\pm.02}$	$0.6401_{\pm.10}$	$0.2839_{\pm.10}$	$0.5351_{\pm.02}$	$0.8335_{\pm.01}$	$0.4867_{\pm.02}$	$0.4730_{\pm.03}$	$0.5638_{\pm.04}$	$0.3828_{\pm.08}$	$0.4084_{\pm.06}$	$0.7194_{\pm.05}$	$0.2731_{\pm.06}$	$0.5335_{\pm.01}$	$0.8231_{\pm.02}$	$0.4489_{\pm.01}$
IP	GCL +GCN	$0.2507_{\pm.06}$	$0.6320_{\pm.03}$	$0.2230_{\pm.03}$	$0.3157_{\pm.02}$	$0.6631_{\pm.04}$	$0.2597_{\pm.02}$	$0.4282_{\pm.02}$	$0.5297_{\pm.05}$	$0.2994_{\pm.07}$	$0.2356_{\pm.04}$	$0.6347_{\pm.03}$	$0.1693_{\pm.06}$	$0.4093_{\pm.01}$	$0.7767_{\pm.01}$	$0.3754_{\pm.01}$
	GCL +FAGCN	$0.3749_{\pm.05}$	$0.7224_{\pm.03}$	$0.3616_{\pm.05}$	$0.4472_{\pm.02}$	$0.7682_{\pm.01}$	$0.4493_{\pm.02}$	$0.4517_{\pm.02}$	$0.5725_{\pm.03}$	$0.4137_{\pm.04}$	$0.4071_{\pm.06}$	$0.7116_{\pm.01}$	$0.2694_{\pm.03}$	$0.5407_{\pm.01}$	$0.8472_{\pm.01}$	$0.5138_{\pm.03}$
	Sim +GCN	$0.2492_{\pm.02}$	$0.5779_{\pm.03}$	$0.1597_{\pm.04}$	$0.2980_{\pm.06}$	$0.6273_{\pm.06}$	$0.2074_{\pm.06}$	$0.3993_{\pm.01}$	$0.5082_{\pm.02}$	$0.2807_{\pm.01}$	$0.2466_{\pm.10}$	$0.6248_{\pm.01}$	$0.1603_{\pm.03}$	$0.4293_{\pm.04}$	$0.7645_{\pm.02}$	$0.3967_{\pm.02}$
	Sim +FAGCN	$0.3763_{\pm.03}$	$0.7246_{\pm.02}$	$0.3561_{\pm.01}$	$0.5161_{\pm.03}$	$0.7984_{\pm.01}$	$0.4625_{\pm.04}$	$0.4386_{\pm.01}$	$0.5547_{\pm.01}$	$0.4018_{\pm.02}$	$0.3983_{\pm.01}$	$0.7118_{\pm.02}$	$0.3020_{\pm.02}$	$0.5411_{\pm.02}$	$0.8549_{\pm.02}$	$0.4955_{\pm.01}$
GCOPE	GCL +GCN	$0.3482_{\pm.07}$	$0.6701_{\pm.05}$	$0.3051_{\pm.07}$	$0.3856_{\pm.04}$	$0.7221_{\pm.04}$	$0.3052_{\pm.06}$	$0.4805_{\pm.04}$	$0.6517_{\pm.04}$	$0.4562_{\pm.03}$	$0.2479_{\pm.01}$	$0.6567_{\pm.00}$	$0.2204_{\pm.01}$	$0.4101_{\pm.03}$	$0.7846_{\pm.01}$	$0.3887_{\pm.03}$
	GCL +FAGCN	$0.3803_{\pm.01}$	$0.7314_{\pm.01}$	$0.3900_{\pm.01}$	$0.5714_{\pm.00}$	$0.8382_{\pm.01}$	$0.5214_{\pm.02}$	$0.4755_{\pm.02}$	$0.5804_{\pm.03}$	$0.4464_{\pm.03}$	$0.4015_{\pm.01}$	$0.7278_{\pm.03}$	$0.2736_{\pm.03}$	$0.5778_{\pm.05}$	$0.8650_{\pm.02}$	$0.5156_{\pm.07}$
	Sim +GCN	$0.3465_{\pm.04}$	$0.6529_{\pm.03}$	$0.2809_{\pm.03}$	$0.3428_{\pm.02}$	$0.6809_{\pm.02}$	$0.3102_{\pm.02}$	$0.3968_{\pm.00}$	$0.5430_{\pm.01}$	$0.3595_{\pm.08}$	$0.2388_{\pm.01}$	$0.6466_{\pm.01}$	$0.2240_{\pm.02}$	$0.4592_{\pm.02}$	$0.8160_{\pm.01}$	$0.4548_{\pm.03}$
	Sim +FAGCN	$0.3867_{\pm.00}$	$0.7345_{\pm.00}$	$0.3774_{\pm.00}$	$0.5645_{\pm.01}$	$0.8457_{\pm.00}$	$0.5169_{\pm.01}$	$0.4654_{\pm.02}$	$0.5676_{\pm.02}$	$0.3913_{\pm.04}$	$0.4079_{\pm.00}$	$0.7356_{\pm.02}$	$0.3070_{\pm.03}$	$0.5511_{\pm.01}$	$0.8642_{\pm.02}$	$0.5332_{\pm.02}$
GEA-CoPe	GCL +GCN	$0.4513_{\pm.02}$	$0.7712_{\pm.01}$	$0.4413_{\pm.01}$	$0.5129_{\pm.06}$	$0.7968_{\pm.02}$	$0.4580_{\pm.06}$	$0.6091_{\pm.04}$	$0.7818_{\pm.02}$	$0.6037_{\pm.04}$	$0.3510_{\pm.08}$	$0.6776_{\pm.01}$	$0.2932_{\pm.01}$	$0.4613_{\pm.04}$	$0.8253_{\pm.02}$	$0.4440_{\pm.03}$
	GCL +FAGCN	$0.4799_{\pm.03}$	$0.7767_{\pm.02}$	$0.4296_{\pm.01}$	$0.5878_{\pm.02}$	$0.8409_{\pm.01}$	$0.5425_{\pm.02}$	$0.4922_{\pm.02}$	$0.5952_{\pm.03}$	$0.4482_{\pm.03}$	$0.3951_{\pm.04}$	$0.6763_{\pm.04}$	$0.2705_{\pm.03}$	$0.6179_{\pm.03}$	$0.8804_{\pm.01}$	$0.5544_{\pm.03}$
	Sim +GCN	$0.4186_{\pm.05}$	$0.7482_{\pm.02}$	$0.4142_{\pm.06}$	$0.5056_{\pm.04}$	$0.7905_{\pm.02}$	$0.4559_{\pm.03}$	$0.5542_{\pm.03}$	$0.7040_{\pm.01}$	$0.5442_{\pm.04}$	$0.3550_{\pm.05}$	$0.6749_{\pm.03}$	$0.3155_{\pm.03}$	$0.4642_{\pm.03}$	$0.8377_{\pm.02}$	$0.4382_{\pm.03}$
	Sim +FAGCN	$0.4526_{\pm.03}$	$0.7717_{\pm.01}$	$0.4364_{\pm.04}$	$0.5990_{\pm.01}$	$0.8546_{\pm.00}$	$0.5605_{\pm.01}$	$0.4975_{\pm.04}$	$0.6966_{\pm.03}$	$0.4799_{\pm.03}$	$0.4427_{\pm.02}$	$0.7363_{\pm.03}$	$0.2956_{\pm.03}$	$0.6156_{\pm.04}$	$0.8727_{\pm.01}$	$0.5199_{\pm.02}$

Table 4. Transfer learning performance (mean±std Acc/AUC/F1) on heterophilic datasets (C-way-1-shot).GCL and Sim respectively represent GraphCL and SimGRACE.

Training schemes	Methods	Wisconsin			Texas			Cornell			Chameleon			Squirrel		
		Acc	AUC	F1	Acc	AUC	F1	Acc	AUC	F1	Acc	AUC	F1	Acc	AUC	F1
Supervised	GCN	$0.4878_{\pm.08}$	$0.7890_{\pm.05}$	$0.4334_{\pm.07}$	$0.6000_{\pm.06}$	$0.6699_{\pm.02}$	$0.4787_{\pm.05}$	$0.3650_{\pm.16}$	$0.5881_{\pm.09}$	$0.2821_{\pm.07}$	$0.2271_{\pm.00}$	$0.5311_{\pm.01}$	$0.1863_{\pm.03}$	$0.2180_{\pm.00}$	$0.5169_{\pm.00}$	$0.1518_{\pm.02}$
	FAGCN	$0.5303_{\pm.06}$	$0.8108_{\pm.04}$	$0.4919_{\pm.09}$	$0.6700_{\pm.04}$	$0.6173_{\pm.05}$	$0.4909_{\pm.08}$	$0.4188_{\pm.17}$	$0.6260_{\pm.08}$	$0.3579_{\pm.11}$	$0.2675_{\pm.02}$	$0.5568_{\pm.00}$	$0.1959_{\pm.01}$	$0.2165_{\pm.00}$	$0.5264_{\pm.00}$	$0.1595_{\pm.03}$
IP	GCL +GCN	$0.5273_{\pm.03}$	$0.7836_{\pm.03}$	$0.4417_{\pm.05}$	$0.6350_{\pm.01}$	$0.6593_{\pm.02}$	$0.4936_{\pm.09}$	$0.3772_{\pm.04}$	$0.6251_{\pm.02}$	$0.3035_{\pm.04}$	$0.2249_{\pm.02}$	$0.5224_{\pm.00}$	$0.1423_{\pm.04}$	$0.2117_{\pm.01}$	$0.5092_{\pm.01}$	$0.1103_{\pm.03}$
	GCL +FAGCN	$0.6049_{\pm.04}$	$0.8362_{\pm.01}$	$0.5588_{\pm.07}$	$0.7433_{\pm.03}$	$0.7038_{\pm.03}$	$0.6141_{\pm.09}$	$0.2688_{\pm.04}$	$0.6267_{\pm.04}$	$0.3642_{\pm.04}$	$0.2412_{\pm.00}$	$0.5470_{\pm.01}$	$0.1845_{\pm.01}$	$0.2143_{\pm.00}$	$0.5086_{\pm.00}$	$0.1728_{\pm.02}$
	Sim +GCN	$0.5058_{\pm.04}$	$0.7749_{\pm.05}$	$0.4610_{\pm.06}$	$0.5938_{\pm.05}$	$0.6425_{\pm.07}$	$0.4257_{\pm.14}$	$0.3638_{\pm.05}$	$0.5852_{\pm.09}$	$0.2768_{\pm.09}$	$0.2237_{\pm.01}$	$0.5293_{\pm.02}$	$0.1569_{\pm.03}$	$0.2063_{\pm.01}$	$0.5103_{\pm.02}$	$0.1550_{\pm.02}$
	Sim +FAGCN	$0.6215_{\pm.02}$	$0.8575_{\pm.00}$	$0.5830_{\pm.04}$	$0.6754_{\pm.12}$	$0.6582_{\pm.02}$	$0.4906_{\pm.04}$	$0.2725_{\pm.06}$	$0.6159_{\pm.04}$	$0.3417_{\pm.04}$	$0.2401_{\pm.01}$	$0.5303_{\pm.00}$	$0.1801_{\pm.00}$	$0.2137_{\pm.00}$	$0.5247_{\pm.00}$	$0.1715_{\pm.01}$
GCOPE	GCL +GCN	$0.5783_{\pm.06}$	$0.8230_{\pm.01}$	$0.4850_{\pm.04}$	$0.6425_{\pm.08}$	$0.6516_{\pm.07}$	$0.5061_{\pm.14}$	$0.3675_{\pm.03}$	$0.6302_{\pm.02}$	$0.2785_{\pm.08}$	$0.2266_{\pm.00}$	$0.5405\ _{\pm.03}$	$0.2092_{\pm.03}$	$0.2205_{\pm.01}$	$0.5256_{\pm.01}$	$0.1713_{\pm.01}$
	GCL +FAGCN	$0.6317_{\pm.04}$	$0.8417_{\pm.01}$	$0.5799_{\pm.06}$	$0.7787_{\pm.03}$	$0.7359_{\pm.01}$	$0.6202_{\pm.05}$	$0.5413_{\pm.06}$	$0.7959_{\pm.02}$	$0.4465_{\pm.01}$	$0.2597_{\pm.01}$	$0.5523_{\pm.01}$	$0.1982_{\pm.03}$	$0.2029_{\pm.00}$	$0.5098_{\pm.00}$	$0.1779_{\pm.01}$
	Sim +GCN	$0.4932_{\pm.08}$	$0.7885_{\pm.05}$	$0.4344_{\pm.07}$	$0.6025_{\pm.13}$	$0.6976_{\pm.01}$	$0.4232_{\pm.11}$	$0.3800_{\pm.02}$	$0.6142_{\pm.03}$	$0.3066_{\pm.05}$	$0.2264_{\pm.00}$	$0.5309\ _{\pm.01}$	$0.1855_{\pm.03}$	$0.2171_{\pm.00}$	$0.5249_{\pm.01}$	$0.1561_{\pm.03}$
	Sim +FAGCN	$0.6670_{\pm.04}$	$0.8684_{\pm.04}$	$0.6287_{\pm.07}$	$0.6800_{\pm.02}$	$0.6677_{\pm.01}$	$0.4850_{\pm.06}$	$0.4200_{\pm.17}$	$0.6265_{\pm.08}$	$0.3582_{\pm.11}$	$0.2786_{\pm.01}$	$0.5589\ _{\pm.02}$	$0.1997_{\pm.02}$	$0.2093\ _{\pm.00}$	$0.5206_{\pm.00}$	$0.1792_{\pm.00}$
GEA-CoPe	GCL +GCN	$0.6000_{\pm.05}$	$0.8210_{\pm.00}$	$0.5885_{\pm.05}$	$0.6590_{\pm.04}$	$0.6591_{\pm.02}$	$0.5788_{\pm.06}$	$0.3812_{\pm.08}$	$0.6344_{\pm.05}$	$0.2848_{\pm.04}$	$0.2371_{\pm.00}$	$0.5440_{\pm.00}$	$0.2028_{\pm.00}$	$0.2464_{\pm.00}$	$0.5474_{\pm.00}$	$0.2203_{\pm.01}$
	GCL +FAGCN	$0.7484_{\pm.01}$	$0.9058_{\pm.00}$	$0.7222_{\pm.01}$	$0.8100_{\pm.03}$	$0.7359_{\pm.01}$	$0.7375_{\pm.05}$	$0.6337_{\pm.01}$	$0.8281_{\pm.02}$	$0.4786_{\pm.01}$	$0.2794_{\pm.02}$	$0.5671_{\pm.02}$	$0.2306_{\pm.01}$	$0.2230_{\pm.00}$	$0.5253_{\pm.00}$	$0.1868_{\pm.00}$
	Sim +GCN	$0.6262_{\pm.04}$	$0.8215_{\pm.01}$	$0.5539_{\pm.04}$	$0.7225_{\pm.05}$	$0.7066_{\pm.01}$	$0.6257_{\pm.06}$	$0.4087_{\pm.08}$	$0.6688_{\pm.03}$	$0.2981_{\pm.03}$	$0.2382_{\pm.02}$	$0.5363_{\pm.02}$	$0.1801_{\pm.01}$	$0.2109_{\pm.00}$	$0.5193_{\pm.01}$	$0.1910_{\pm.00}$
	Sim +FAGCN	$0.7774_{\pm.00}$	$0.9243_{\pm.01}$	$0.7469_{\pm.01}$	$0.7475_{\pm.00}$	$0.6810_{\pm.00}$	$0.5957_{\pm.03}$	$0.5237_{\pm.06}$	$0.7996_{\pm.03}$	$0.3814_{\pm.05}$	$0.2407_{\pm.02}$	$0.5324_{\pm.01}$	$0.1993_{\pm.01}$	$0.2204_{\pm.00}$	$0.5342_{\pm.00}$	$0.2073_{\pm.01}$

-the-art methods (MDGPT [49], MDGFM [50], SAMGPT [51] and UniPrompt [52]) were selected for comparison. As shown in Table 5. Compared to MDGPT, which employs domain tokens for explicit feature semantic alignment, our model implicitly enhances the discriminative power and domain invariance of features by introducing contrastive-learning signals during the normalization process, thereby avoiding potential semantic bias caused by explicit token alignment. While MDGFM relies on complex graph-structure learning for explicit topological reconstruction, our model utilizes the more lightweight ContraNorm to implicitly improve robustness, maintaining efficiency while avoiding the significant overhead and potential structural distortion risks associated with graph topology-aware alignment. Unlike SAMGPT, which depends on structural tokens for layer-wise topological alignment, our model achieves dynamic, attention-weighted fusion of multi-source domain contributions *via* GEANet within the coordinator, eliminating the need for introducing fixed structural parameters. In contrast to the general prompt framework UniPrompt, our model is specifically designed for cross-domain graph learning. The dynamic domain-adaptation capability provided by GEANet is significantly superior to UniPrompt's static task templates. Simultaneously, the powerful function-approximation capability of the KAN classifier head far exceeds that of commonly used linear or shallow classifiers in few-shot scenarios.

Overall, through the synergistic design of "dynamic fusion, contrastive enhancement and strong-fitting classification," our model demonstrates excellent performance across three key aspects: adaptive integration of multi-domain knowledge, robustness of representations and adaptation to downstream tasks. The best reported node-classification performance of these methods across ten datasets was compared with the best performance achieved by our proposed framework. As shown in Table 6, the best performance of our proposed framework clearly surpasses that of the other methods, demonstrating its effectiveness.

Table 5. Cross-domain graph methods.

Method	Core Architecture	Alignment Mechanism	Domain Adaptation	Classification Head
MDGPT [48]	Domain Tokens +Dual Prompts	Domain Token Explicit Alignment	Unified Prompt +Mixed Prompt	Linear Classifier or Prototypical Classifier
MDGFM [49]	Graph Structure Learning +Dual Prompts	Graph Structure Learning Explicit Alignment	Meta Prompt +Task Prompt	Prototypical Classifier or Linear Classifier
SAMGPT [50]	Structure Tokens +Dual Prompts	Structure Token Explicit Alignment	Global Prompt +Specific Prompt	Prototypical Classifier
UniPrompt [51]	Unified Task Template+Learnable Prompts	Task Template Alignment	General Prompting	Linear Classifier or Shallow MLP
GEA-CoPe	External Attention Enhanced Coordinator	Coordinator Implicit Semantic Alignment	Coordinator Adaptive Weighting Prompt /Fine-tuning	KAN: Strong Nonlinear Function Approximation

4.3 Reconstruction Loss Analysis

On the Citeseer dataset, the proposed method was systematically evaluated for its impact on downstream node classification tasks under different reconstruction loss coefficients, with a comparative analysis of supervised-learning methods used to assess the effectiveness of the reconstruction module. In the specific experimental setup, FAGCN was adopted as the backbone network architecture and GraphCL was employed as the graph contrastive-learning pre-training strategy. To ensure a fair comparison, all other hyper-parameter configurations were kept identical across the compared methods. A detailed comparison of the experimental results is shown in Figure 6, with a comprehensive analysis conducted based on three evaluation metrics: node-classification accuracy (Acc), area under the ROC curve (AUC) and F1-score.

Based on the experimental data, the following conclusions can be drawn: First, without the reconstruction module ($\lambda = 0.0$), the framework already outperforms supervised pre-training, demonstrating the effectiveness of the coordinator design. Second, when the reconstruction module is introduced and λ is set to 0.2, the model achieves optimal performance, surpassing not only supervised pre-training, but also the framework without reconstruction ($\lambda = 0.0$). This improvement benefits from the reconstruction module's ability to align graph features across datasets, enabling the graph neural network to more effectively learn common information from multi-source cross-domain data. However, when λ exceeds 0.2, model performance begins to decline, eventually falling below both supervised pre-training and the performance without reconstruction. This is attributed to excessively large λ values causing the model to over-prioritize the reconstruction task, thereby weakening the learning effectiveness of the primary pre-training task. In summary, introducing the reconstruction module with a relatively small λ value is a key factor in ensuring the effectiveness of the framework method.

Table 6. Node-classification accuracy for cross-domain graph pre-training methods.

Methods	Cora	Citeseer	Pubmed	Computers	Photos	Wisconsin	Texas	Cornell	Chameleon	Squirrel
MDGPT [48]	$0.4226_{\pm.10}$	$0.4240_{\pm.09}$	$0.4982_{\pm.08}$	$0.4216_{\pm.11}$	$0.5496_{\pm.10}$	$0.5040_{\pm.15}$	$0.5976_{\pm.12}$	$0.5419_{\pm.13}$	$0.2804_{\pm.04}$	$0.2441_{\pm.07}$
MDGFM [49]	$0.4483_{\pm.07}$	$0.4218_{\pm.06}$	$0.4684_{\pm.07}$	-	-	-	-	$0.4077_{\pm.05}$	$0.2836_{\pm.03}$	$0.2430_{\pm.03}$
SAMGPT [50]	$0.4680_{\pm.11}$	$0.3638_{\pm.09}$	$0.5025_{\pm.10}$	$0.4522_{\pm.08}$	$0.5871_{\pm.08}$	$0.5229_{\pm.14}$	$0.6679_{\pm.10}$	$0.5934_{\pm.09}$	$0.2812_{\pm.08}$	$0.2475_{\pm.06}$
UniPrompt[51]	$0.4537_{\pm.09}$	$0.4325_{\pm.09}$	$0.5501_{\pm.03}$	-	-	-	-	$0.5158_{\pm.09}$	$0.2514_{\pm.05}$	$0.2429_{\pm.03}$
GEA-CoPe	$0.4799_{\pm.03}$	$0.5990_{\pm.01}$	$0.6091_{\pm.04}$	$0.4427_{\pm.02}$	$0.6179_{\pm.03}$	$0.7774_{\pm.00}$	$0.8100_{\pm.03}$	$0.6337_{\pm.01}$	$0.2794_{\pm.00}$	$0.2464_{\pm.00}$

* “——” denotes that the official code has not been released for implementation on these datasets.

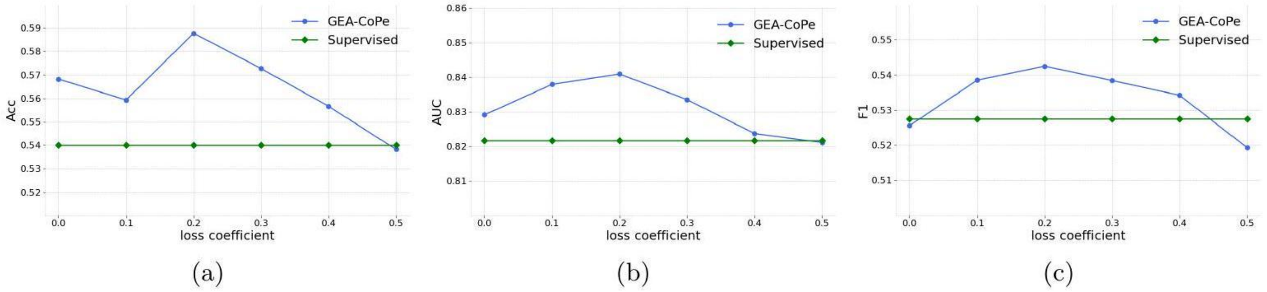


Figure 6. Node-classification performance of GEA-CoPe on Citeseer under C-way-1-shot setting. (a)Variation of Acc with reconstruction loss coefficient. (b)Variation of AUC with reconstruction loss coefficient. (c)Variation of F1-score with reconstruction loss coefficient.

4.4 Transferring by Graph Prompt

To transfer and apply knowledge learned from upstream tasks to downstream tasks, two methods are selected: fine-tuning and graph-prompting techniques. Next, the feasibility of knowledge transfer *via* graph prompting techniques is tested. More specifically, ProG [17] method is adopted, which is a revolutionary graph neural network transfer-learning paradigm. It constructs a lightweight, learnable "prompt graph" relevant to the downstream task and structurally integrates this prompt graph with the original input graph, thereby effectively "prompting" the frozen pre-trained GNN model with task information.

The downstream datasets Cora, citeseer, Wisconsin and Texas were selected for the node classification task, including two homophilic and two heterophilic datasets, to evaluate model performance. The experimental results are shown in Table 7. To rigorously and intuitively assess the viability of the ProG method, the results were compared with the results of supervised methods and the results of GEA-CoPe using fine-tuning.

By comparing the experimental results, the following conclusions can be drawn: GEA-CoPe demonstrates superior performance compared to other methods, regardless of whether knowledge is transferred using fine-tuning or the ProG method. Particularly in the node-classification task, GEA-CoPe utilizing ProG achieves positive transfer with the fewest tunable parameters. However, the model using ProG performs slightly worse than the model using fine-tuning. Models employing these two methods generally outperform supervised methods. Through analysis of the results, it can be concluded that our proposed framework is favorable for prompt learning on downstream tasks.

4.5 Impact of Attention Heads

To investigate the impact of the number of external attention heads on GEA-CoPe, the performance of GEA-CoPe method with varying numbers of attention heads in downstream node-classification task was

Table 7. Cross-domain transfer-learning performance (mean ± std Acc/AUC/F1) of GEA-CoPe with ProG (C-way-1-shot). GCL and Sim, respectively, representing GraphCL and SimGRACE.

Training schemes	Methods	Cora			Pubmed		
		Acc	AUC	F1	Acc	AUC	F1
Supervised	FAGCN	$0.3359_{\pm.02}$	$0.6401_{\pm.10}$	$0.2839_{\pm.10}$	$0.4730_{\pm.03}$	$0.5638_{\pm.04}$	$0.3828_{\pm.08}$
GEA-CoPe +ProG	GCL-FAGCN	$0.3419_{\pm.01}$	$0.7230_{\pm.02}$	$0.3041_{\pm.08}$	$0.4750_{\pm.02}$	$0.6732_{\pm.04}$	$0.4205_{\pm.01}$
	Sim-FAGCN	$0.4015_{\pm.02}$	$0.7265_{\pm.01}$	$0.3700_{\pm.03}$	$0.4450_{\pm.00}$	$0.5922_{\pm.01}$	$0.4384_{\pm.01}$
GEA-CoPe +finetuning	GCL-FAGCN	$0.4699_{\pm.03}$	$0.7767_{\pm.02}$	$0.4296_{\pm.01}$	$0.4922_{\pm.02}$	$0.5952_{\pm.03}$	$0.4482_{\pm.03}$
	Sim-FAGCN	$0.4526_{\pm.03}$	$0.7717_{\pm.01}$	$0.4364_{\pm.04}$	$0.4975_{\pm.04}$	$0.6966_{\pm.03}$	$0.4799_{\pm.03}$
Training schemes	Methods	Wisconsin			Texas		
		Acc	AUC	F1	Acc	AUC	F1
Supervised	FAGCN	$0.5303_{\pm.06}$	$0.8108_{\pm.04}$	$0.4919_{\pm.09}$	$0.6700_{\pm.04}$	$0.6173_{\pm.05}$	$0.4909_{\pm.08}$
GEA-CoPe +ProG	GCL-FAGCN	$0.5467_{\pm.00}$	$0.8216_{\pm.00}$	$0.4863_{\pm.02}$	$0.7712_{\pm.03}$	$0.6847_{\pm.00}$	$0.6412_{\pm.07}$
	Sim-FAGCN	$0.7394_{\pm.00}$	$0.8944_{\pm.01}$	$0.6982_{\pm.02}$	$0.7400_{\pm.03}$	$0.6645_{\pm.00}$	$0.6420_{\pm.07}$
GEA-CoPe +finetuning	GCL-FAGCN	$0.7484_{\pm.01}$	$0.9058_{\pm.00}$	$0.7222_{\pm.01}$	$0.8104_{\pm.03}$	$0.7359_{\pm.01}$	$0.7375_{\pm.05}$
	Sim-FAGCN	$0.7774_{\pm.00}$	$0.9243_{\pm.01}$	$0.7469_{\pm.01}$	$0.7475_{\pm.00}$	$0.6810_{\pm.00}$	$0.5957_{\pm.03}$

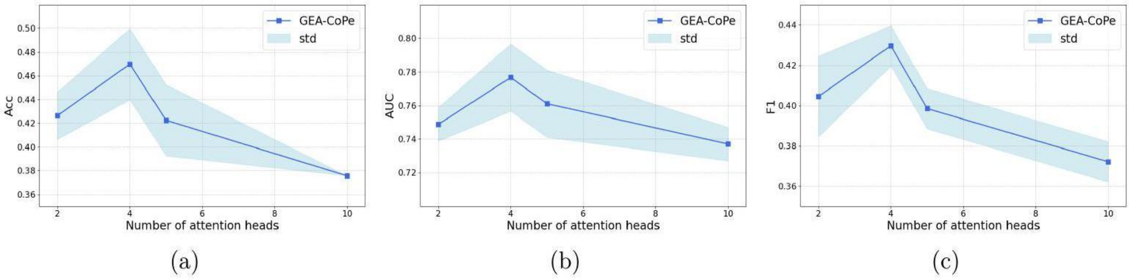


Figure 7. Node-classification performance (mean ± std) of GEA-CoPe on Cora under C-way-1shot setting. (a)Variation of Acc with the number of attention heads. (b)Variation of AUC with the number of attention heads. (c)Variation of F1-score with the number of attention heads.

compared. Specifically, FAGCN is selected as the backbone model, GraphCL is selected as the pre-training strategy, all other super-parameters are consistent and the node-classification task is performed on the Cora dataset. The experimental results, presented in Figure 7, primarily demonstrate the Acc, AUC and F1-score metrics.

From the figure, it can be observed that performance initially increases and then decreases with the growing number of attention heads: An insufficient number of attention heads leads to lower performance due to insufficient representation capacity. Increasing the number of heads enables the model to capture richer neighborhood information, significantly improving the Acc, AUC and F1-score. However, when the number of attention heads becomes excessive, performance declines as the model suffers from over-fitting or noise interference. When the external number of attention heads is 4, all metrics reach their peaks, resulting in the best node-classification performance.

4.6 Analysis of Neural Network Layers

To systematically evaluate the impact of neural-network depth on model performance, a controlled variable experiment was designed. While keeping the hidden-layer dimensionality and other hyper-parameters fixed, the number of graph neural-network layers was progressively increased. Experiments were conducted uniformly using the GraphCL and FAGCN methods on the Photos and Texas datasets to assess the influence of GNN depth on framework performance, with evaluation metrics including Accuracy, AUC and F1-score. The results are shown in Figure 8.

The experimental results clearly demonstrate that as the number of layers increases, the model performance shows an upward trend. When the number of layers is less than 8, the node-classification

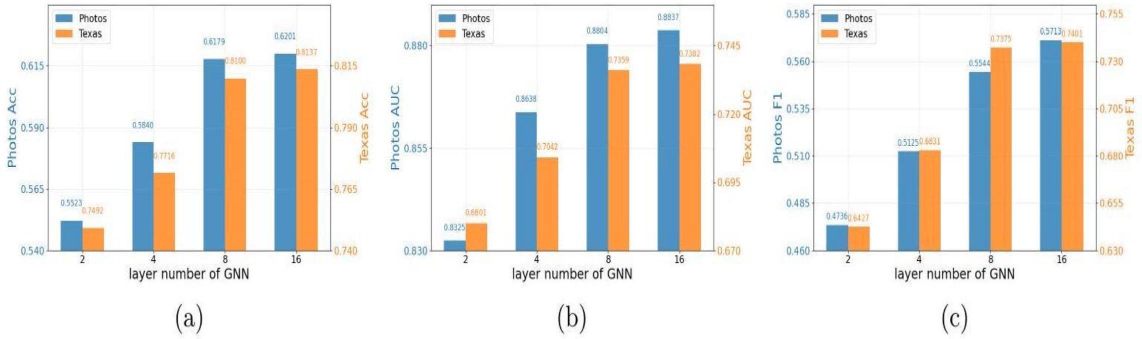


Figure 8. Node-classification performance of GEA-CoPe on Photos and Texas under C-way-1shot setting. (a) The variation of Acc with the number of layers in GNNs. (b) The variation of AUC with the number of layers in GNNs. (c) The variation of F1-score with the number of layers in GNNs.

performance increases markedly, whereas beyond 8 layers, the improvement becomes more gradual. This occurs, because excessively deep network structures are prone to issues, such as gradient vanishing or over-smoothing during propagation, which can impair the model's ability to discriminate local node features. While too few layers may lead to under-fitting, too many layers significantly increase computational time. Selecting 8 GNN layers achieves an optimal balance between node-classification accuracy and runtime.

4.7 Ablation Experiments

To thoroughly investigate the impact of individual components in GEA-CoPe on the overall model performance, multiple ablation studies were conducted, analyzing the effects of the graph external-attention mechanism, dual contrast normalization and the KAN classification head. Under the unified framework employing both SimGRACE and FAGCN, four variants were constructed: Variant 1 incorporates only graph external attention, excludes dual contrast normalization and uses a traditional classification head; Variant 2 removes the external attention from the base model; Variant 3 utilizes traditional graph neural networks for data processing; Variant 4 adopts a traditional classification head. Experiments were performed on the base model and all variants, as shown in Table 8 and Table 9.

Table 8. Node-classification performance on homophilic datasets (C-way-1-shot).

Methods	Cora			Pubmed			Photos		
	Acc	AUC	F1	Acc	AUC	F1	Acc	AUC	F1
Base Model	$0.3867_{\pm.00}$	$0.7345_{\pm.00}$	$0.3774_{\pm.02}$	$0.4654_{\pm.02}$	$0.5676_{\pm.02}$	$0.3913_{\pm.04}$	$0.5541_{\pm.01}$	$0.8342_{\pm.02}$	$0.5012_{\pm.02}$
Variant 1	$0.4023_{\pm.02}$	$0.7419_{\pm.03}$	$0.3948_{\pm.01}$	$0.4763_{\pm.03}$	$0.6192_{\pm.01}$	$0.4358_{\pm.02}$	$0.5924_{\pm.03}$	$0.8671_{\pm.00}$	$0.5136_{\pm.01}$
Variant 2	$0.4186_{\pm.02}$	$0.7463_{\pm.02}$	$0.4292_{\pm.02}$	$0.4792_{\pm.02}$	$0.6271_{\pm.04}$	$0.4030_{\pm.06}$	$0.6030_{\pm.02}$	$0.8460_{\pm.03}$	$0.5163_{\pm.02}$
Variant 3	$0.4072_{\pm.03}$	$0.7128_{\pm.02}$	$0.4037_{\pm.03}$	$0.4629_{\pm.00}$	$0.5716_{\pm.04}$	$0.4559_{\pm.01}$	$0.5943_{\pm.02}$	$0.8654_{\pm.01}$	$0.5276_{\pm.03}$
Variant 4	$0.4012_{\pm.01}$	$0.7427_{\pm.01}$	$0.4009_{\pm.01}$	$0.4886_{\pm.01}$	$0.6401_{\pm.02}$	$0.4689_{\pm.01}$	$0.6120_{\pm.03}$	$0.8686_{\pm.01}$	$0.5047_{\pm.03}$
GEA-CoPe	$0.4526_{\pm.03}$	$0.7717_{\pm.01}$	$0.4364_{\pm.04}$	$0.4975_{\pm.04}$	$0.6966_{\pm.03}$	$0.4799_{\pm.03}$	$0.6156_{\pm.04}$	$0.8727_{\pm.01}$	$0.5199_{\pm.02}$

As evidenced by the table, the base model performs the worst across all datasets, while the variant methods exhibit certain advantages in specific scenarios, but demonstrate inconsistent performance. GEACoPe achieves particularly marked improvements on heterophilic datasets, indicating its effectiveness in handling class-distribution imbalance and complex connectivity patterns. The proposed framework outperforms both the base model and the variants in the vast majority of cases, highlighting its comprehensive superiority, especially on heterophilic datasets where it shows significant enhancements. This demonstrates the framework's strong generalization capability in effectively addressing node-classification tasks across diverse graph structures.

4.8 Robustness Analysis

To evaluate the robustness of the model, three typical types of feature perturbation-Gaussian noise injection, feature sparsification and node-feature masking - were introduced to simulate common data-

Table 9. Node-classification performance on heterophilic datasets (C -way-1-shot).

Methods	Wisconsin			Texas			Squirrel		
	Acc	AUC	F1	Acc	AUC	F1	Acc	AUC	F1
Base Model	$0.6070_{\pm.04}$	$0.8284_{\pm.04}$	$0.5287_{\pm.07}$	$0.6800_{\pm.02}$	$0.6477_{\pm.01}$	$0.4850_{\pm.06}$	$0.2093_{\pm.00}$	$0.5106_{\pm.00}$	$0.1692_{\pm.01}$
Variant 1	$0.7153_{\pm.02}$	$0.8914_{\pm.03}$	$0.7140_{\pm.02}$	$0.7092_{\pm.05}$	$0.6526_{\pm.03}$	$0.5413_{\pm.02}$	$0.2146_{\pm.01}$	$0.5283_{\pm.01}$	$0.1892_{\pm.00}$
Variant 2	$0.7285_{\pm.01}$	$0.8987_{\pm.01}$	$0.7206_{\pm.01}$	$0.6525_{\pm.17}$	$0.6739_{\pm.05}$	$0.5308_{\pm.13}$	$0.2133_{\pm.00}$	$0.5219_{\pm.00}$	$0.1751_{\pm.01}$
Variant 3	$0.6100_{\pm.04}$	$0.8317_{\pm.00}$	$0.5398_{\pm.05}$	$0.7125_{\pm.06}$	$0.6509_{\pm.04}$	$0.4995_{\pm.02}$	$0.2174_{\pm.00}$	$0.5279_{\pm.00}$	$0.1604_{\pm.01}$
Variant 4	$0.5919_{\pm.02}$	$0.8278_{\pm.04}$	$0.4861_{\pm.10}$	$0.7300_{\pm.03}$	$0.6686_{\pm.03}$	$0.5738_{\pm.05}$	$0.2142_{\pm.00}$	$0.5265_{\pm.00}$	$0.1719_{\pm.02}$
GEA-CoPe	$0.7774_{\pm.00}$	$0.9243_{\pm.01}$	$0.7469_{\pm.01}$	$0.7475_{\pm.00}$	$0.6810_{\pm.00}$	$0.5957_{\pm.03}$	$0.2197_{\pm.00}$	$0.5302_{\pm.00}$	$0.1806_{\pm.01}$

quality issues in real-world applications, such as noise, sparse node features or partially missing attributes. The experiments were conducted on both homophilic and heterophilic datasets as target domains under a 1-shot learning setting. Methods including GraphCL and FAGCN were employed, with pre-training performed on the remaining nine datasets and downstream tasks carried out on the target dataset. Perturbations of the same type and intensity were applied in both stages to comprehensively assess the model's robustness under impaired feature conditions.

Table 10. Node-classification performance with Gaussian noise on GEA-CoPe (C-way-1-shot).

Standard deviation	Cora			Texas		
	Acc	AUC	F1	Acc	AUC	F1
0.0	$0.4799_{\pm.03}$	$0.7767_{\pm.02}$	$0.4296_{\pm.01}$	$0.8100_{\pm.03}$	$0.7359_{\pm.01}$	$0.7375_{\pm.05}$
0.3	$0.4648_{\pm.01}$	$0.7628_{\pm.02}$	$0.4055_{\pm.01}$	$0.7875_{\pm.05}$	$0.7047_{\pm.03}$	$0.6563_{\pm.09}$
0.5	$0.4512_{\pm.03}$	$0.7514_{\pm.03}$	$0.3921_{\pm.03}$	$0.7650_{\pm.06}$	$0.6925_{\pm.04}$	$0.6314_{\pm.11}$
0.7	$0.4326_{\pm.04}$	$0.7398_{\pm.04}$	$0.3787_{\pm.04}$	$0.7412_{\pm.07}$	$0.6783_{\pm.05}$	$0.6059_{\pm.12}$

The experimental results are shown in Tables 10, 11 and 12. Overall, the model demonstrates notable robustness and superiority when facing various feature perturbations. Under Gaussian-noise interference, the model achieves cross-graph feature smoothing through its coordinator. Even under high-intensity noise, it maintains high accuracy, indicating its strong filtering capability against random errors. In the feature-sparsification experiments, when 90% of features are zeroed out, the model exhibits only a slight drop in accuracy, benefiting from the cross-graph information compensation and structural enhancement enabled by the coordinator and external-attention mechanism. Particularly in heterophilic graphs, the rich topological structure provides critical information compensation, resulting in significantly better performance retention compared to homophilic graphs. In the most challenging scenario of node-feature masking, where 70% of node features are completely absent, the accuracy on the Texas dataset remains at 72.64%. This suggests that the model does not simply rely on complete feature inputs, but can effectively capture key patterns from partially masked features and integrate graph structural information for reliable inference. Comprehensive analysis indicates that the model's robustness stems from its dynamic adaptive mechanism: the external attention and coordinator can intelligently adjust the weights of intra- and inter-graph information flow according to the type and intensity of interference, achieving synergy among feature smoothing, missing feature compensation and structural enhancement. Moreover, multi-graph pre-training endows the model with more generalized feature invariance. This enables the model not only to perform well under ideal data conditions, but also to maintain stable performance with low-quality and incomplete feature data commonly encountered in real-world scenarios.

Table 11. Node-classification performance with feature sparsification on GEA-CoPe (C-way-1shot).

Sparse scale	Cora			Texas		
	Acc	AUC	F1	Acc	AUC	F1
0%	$0.4799_{\pm.03}$	$0.7767_{\pm.02}$	$0.4296_{\pm.01}$	$0.8100_{\pm.03}$	$0.7359_{\pm.01}$	$0.7375_{\pm.05}$
50%	$0.4646_{\pm.02}$	$0.7678_{\pm.01}$	$0.4181_{\pm.03}$	$0.7737_{\pm.04}$	$0.7252_{\pm.02}$	$0.6485_{\pm.09}$
70%	$0.4482_{\pm.03}$	$0.7543_{\pm.03}$	$0.4027_{\pm.04}$	$0.7419_{\pm.05}$	$0.7128_{\pm.03}$	$0.6184_{\pm.10}$
90%	$0.4185_{\pm.04}$	$0.7326_{\pm.04}$	$0.3789_{\pm.05}$	$0.7013_{\pm.07}$	$0.6934_{\pm.04}$	$0.5742_{\pm.12}$

Table 12. Node classification performance with node feature masking on GEA-CoPe (C-way-1shot).

Mask scale	Cora			Texas		
	Acc	AUC	F1	Acc	AUC	F1
0%	$0.4799_{\pm.03}$	$0.7767_{\pm.02}$	$0.4296_{\pm.01}$	$0.8100_{\pm.03}$	$0.7359_{\pm.01}$	$0.7375_{\pm.05}$
30%	$0.4324_{\pm.02}$	$0.7716_{\pm.03}$	$0.4094_{\pm.04}$	$0.7925_{\pm.04}$	$0.7263_{\pm.01}$	$0.7278_{\pm.07}$
50%	$0.4018_{\pm.03}$	$0.7582_{\pm.04}$	$0.3876_{\pm.05}$	$0.7637_{\pm.05}$	$0.7129_{\pm.02}$	$0.6843_{\pm.09}$
70%	$0.3685_{\pm.04}$	$0.7427_{\pm.05}$	$0.3621_{\pm.06}$	$0.7264_{\pm.06}$	$0.6985_{\pm.03}$	$0.6328_{\pm.10}$

5. CONCLUSION

This study addresses the negative-transfer problem in cross-domain graph pre-training under few-shot learning scenarios by proposing a novel multi-component framework named GEA-CoPe. The inherent structural and semantic discrepancies between graph domains significantly hinder effective knowledge transfer, while existing methods often fail to resolve this issue due to their limited adaptability and lack of explicit constraints on feature consistency. The proposed framework innovatively integrates multi-head external attention with a graph coordinator, enabling dynamic and adaptive cross-graph semantic alignment to bridge domain gaps while preserving unique structural information. The introduced dual feature-normalization strategy, which combines intra-layer node-similarity constraints with a cross-layer distribution-alignment loss, effectively mitigates feature drift and enhances the robustness and stability of pre-trained representations. Furthermore, by incorporating Kolmogorov-Arnold Networks (KAN) with parameter-adaptive activation functions, the model gains superior non-linear representation capability and improved interpretability, allowing it to better capture complex topological dependencies. Extensive experiments conducted on ten real-world graph datasets demonstrate that GEA-CoPe significantly outperforms existing methods in both cross-domain generalization and few-shot node-classification tasks. The model's ability to focus on critical graph structures while maintaining consistent feature distributions throughout propagation highlights its practical potential in complex and resource-constrained environments.

Despite the encouraging results, the proposed framework has certain limitations. Its performance still partially depends on the quality and diversity of the pre-training data. Moreover, the increased model complexity may require additional computational resources during training. Future work will focus on extending the framework to handle more dynamic and heterogeneous graph structures, optimizing its efficiency for large-scale deployment and exploring its integration with other advanced pre-training paradigms.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (Grant No. 52174184).

REFERENCES

[1] D. Bhattacharjee et al., "Vision Transformer Adapters for Generalizable Multitask Learning," Proc. of the IEEE/CVF Int. Conf. on Computer Vision (ICCV), pp. 19015-19026, Paris, France, 2023.

[2] M. Sun et al., "GPPT: Graph Pre-training and Prompt Tuning to Generalize Graph Neural Networks," Proc. of the 28th ACM SIGKDD Conf. on Knowledge Discovery and Data Mining (KDD), pp. 1717-1727, DOI: 10.1145/3534678.3539249 2022.

[3] J. Liu et al., "Graph Foundation Models: Concepts, Opportunities and Challenges," IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), Vol. 47, no. 6, pp. 5023-5044, 2025.

[4] X. Sun et al., "All in One: Multi-task Prompting for Graph Neural Networks," Proc. of the 29th ACM SIGKDD Conf. on Knowledge Discovery and Data Mining (KDD), pp. 2120–2131, DOI: 10.1145/3580305.3599256, 2023.

[5] A. A. Khan et al., "Blockchain-enabled Secure Internet of Medical Things (IoMT) Architecture for Multi-Modal Data Fusion in Precision Cancer Diagnosis and Continuous Monitoring," Journal of Cloud Computing, vol. 14, p. 58, 2025.

[6] B.-S. Shi et al., "Domain Adaptation for Graph Representation Learning: Challenges, Progress and Prospects," Journal of Computer Science and Technology, vol. 40, pp. 283–300, 2025.

[7] Y. Xue et al., "A Review on Transferability Estimation in Deep Transfer Learning," IEEE Transactions on Artificial Intelligence (IEEE TAIS), vol. 5, no.12, pp. 5894 - 5914, 2024.

[8] A. A. Laghari et al., "A Novel and Secure Artificial Intelligence Enabled Zero Trust Intrusion Detection

in Industrial Internet of Things Architecture," Scientific Reports, vol. 15, p. 26843, 2025.

[9] X. Wu et al., "ProCom: A Few-shot Targeted Community Detection Algorithm," Proc. of the 30th ACM SIGKDD Conf. on Knowledge Discovery and Data Mining (KDD), pp. 3414–3424, DOI: 10.1145/3637528.3671749, 2024.

[10] L. Sun et al., "RiemannGFM: Learning a Graph Foundation Model from Riemannian Geometry," Proc. of the ACM on Web Conf. (WWW), pp. 1154–1165, DOI: 10.1145/3696410.3714952, 2025.

[11] Z. Wang et al., "Negative as Positive: Enhancing Out-of-distribution Generalization for Graph Contrastive Learning," Proc. of the 47th Int. ACM SIGIR Conf. on Research and Development in Information Retrieval (SIGIR), pp. 2548–2552, DOI: 10.1145/3626772.3657927, 2024.

[12] Q. Chen et al., "DAGPrompt: Pushing the Limits of Graph Prompting with a Distribution-aware Graph Prompt Tuning Approach," Proc. of the ACM on Web Conf. (WWW), pp. 4346–4358, DOI: 10.1145/3696410.3714917, 2025.

[13] X. Huang et al., "Enhancing Cross-domain Link Prediction *via* Evolution Process Modeling," Proc. of the ACM on Web Conf. (WWW), pp. 2158–2171, DOI: 10.1145/3696410.3714792, 2025.

[14] L. Kong et al., "Gofa: A Generative One-for-all Model for Joint Graph Language Modeling," Proc. of the 13th Int. Conf. on Learning Representations (ICLR), DOI:10.1021/acsengineeringau.3c00058.s001, 2025.

[15] M. Zhang et al., "GraphTranslator: Aligning Graph Model to Large Language Model for Open-ended Tasks," Proc. of the ACM Web Conf. (WWW), pp. 1003–1014, DOI: 10.1145/3589334.3645682, 2024.

[16] Y. You et al., "Graph Contrastive Learning with Augmentations," Advances in Neural Information Processing Systems (NeurIPS), vol. 33, pp. 5812–5823, 2020.

[17] J. Xia et al., "SimGRACE: A Simple Framework for Graph Contrastive Learning without Data Augmentation," Proc. of the ACM Web Conf. (WWW), pp. 1070–1079, DOI: 10.1145/3485447.3512156, 2022.

[18] Z. Hu et al., "GPT-GNN: Generative Pre-training of Graph Neural Networks," Proc. of the 26th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD), pp. 1857–1867, DOI: 10.1145/3394486.3403237, 2020.

[19] Z. Hou et al., "GraphMAE: Self-supervised Masked Graph Auto-encoders," Proc. of the 28th ACM SIGKDD Conf. on Knowledge Discovery and Data Mining (KDD), pp. 594–604, DOI: 10.1145/3534678.3539321, 2022.

[20] H. Yan et al., "Hierarchical Graph Contrastive Learning," Proc. of the Joint European Conf. on Machine Learning and Knowledge Discovery in Databases (ECML-PKDD), vol. 14170, pp. 700–715, 2023.

[21] Q. Dai et al., "Graph Transfer Learning *via* Adversarial Domain Adaptation with Graph Convolution," IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 5, pp. 4908–4922, 2022.

[22] T. Vayer et al., "Fused Gromov-Wasserstein Distance for Structured Objects," Algorithms, vol. 13, no. 9, p. 212, 2020.

[23] Z. Hu, Y. Dong, K. Wang and Y. Sun, "Heterogeneous Graph Transformer," Proc. of the Web Conf. (WWW), pp. 2704–2710, DOI: 10.1145/3366423.3380027, 2020.

[24] G. Wan et al., "Reinforcement Learning-based Meta-path Discovery in Large-scale Heterogeneous Information Networks," Proc. of the AAAI Conf. on Artifi. Intell., vol. 34, no. 04, pp. 6094–6101, 2020.

[25] A. Pareja et al., "EvolveGCN: Evolving Graph Convolutional Networks for Dynamic Graphs," Proc. of the AAAI Conf. on Artificial Intelligence, vol. 34, no. 04, pp. 5363–5370, 2020.

[26] R. Trivedi et al., "DyRep: Learning Representations over Dynamic Graphs," Proc. of Int. Conf. on Learning Representations (ICLR), DOI:10.32920/26883523.v1, 2019.

[27] G. Song, Y. Zhang, L. Xu and H. Lu, "Domain Adaptive Network Embedding," IEEE Transactions on Big Data, vol. 8, no. 5, pp. 1220–1232, 2020.

[28] L. Chen et al., "Graph Optimal Transport for Cross-Domain Alignment," Proc. of the 37th Int. Conf. on Machine Learning (ICML), vol.119, pp. 1542–1553, 2020.

[29] J. Liang, M. Chen and J. Liang, "Graph External Attention Enhanced Transformer," Proc. of Int. Conf. on Machine Learning (ICML), Vol. 235 pp. 29560–29574, 2024.

[30] W. Jin et al., "Self-supervised Learning on Graphs: Deep Insights and New Direction," [Online], Available: https://doi.org/10.48550/arXiv.2006.10141, 2020.

[31] X. Guo et al., "ContraNorm: A Contrastive Learning Perspective on Over-smoothing and Beyond," Proc. of the 11th Int. Conf. on Learning Representations (ICLR), DOI: 10.48550/arXiv.2303.06562, 2023.

[32] Z. Liu et al., "KAN: Kolmogorov-Arnold Networks," Proc. of the 13th Int. Conf. on Learning Representations (ICLR), DOI: 10.31224/5413, 2025.

[33] R. Rossi and N. Ahmed, "The Network Data Repository with Interactive Graph Analytics and Visualization," Proc. of the 29th AAAI Conf. on Artificial Intelligence, vol. 29, no. 1, DOI: 10.1609/aaai.v29i1.9277, 2015.

[34] P. Sen et al., "Collective Classification in Network Data," AI Magazine, vol. 29, no. 3, p. 93, 2008.

[35] G. Namata et al., "Query-driven Active Surveying for Collective Classification," Proc. of the 10th Int. Workshop on Mining and Learning with Graphs (MLG), vol. 8, pp. 1-8, Edinburgh, UK, 2012.

[36] J. McAuley et al., "Image-based Recommendations on Styles and Substitutes," Proc. of the 38th Int. ACM

SIGIR Conf. on Research and Development in Information Retrieval (SIGIR), pp. 43–52, DOI: 10.1145/2766462.2767755, 2015.

[37] O. Shchur, M. Mumme, A. Bojchevski and S. Günnemann, "Pitfalls of Graph Neural Network Evaluation," [Online], Available: https://doi.org/10.48550/arXiv.1811.05868, 2018.

[38] H. Pei et al., "Geom-GCN: Geometric Graph Convolutional Networks," Proc. of the Int. Conf. on Learning Representations (ICLR), DOI:10.48550/arXiv.2002.05287, 2020.

[39] Z. Xu et al., "Node Classification Beyond Homophily: Towards a General Solution," Proc. of the 29th ACM SIGKDD Conf. on Knowledge Discovery and Data Mining (KDD), pp. 2862–2873, 2023.

[40] S. Luan et al., "When Do Graph Neural Networks Help with Node Classification: Investigating the Homophily Principle on Node Distinguishability," NeurIPS J., vol. 36, pp. 28748–28760, 2023.

[41] T. N. Kipf and M. Welling, "Semi-supervised Classification with Graph Convolutional Networks," Proc. of the Int. Conf. on Learning Representations (ICLR), DOI: 10.18178/wcse.2019.06.016, 2016.

[42] D. Bo, X. Wang, C. Shi and H. Shen, "Beyond Low-frequency Information in Graph Convolutional Networks," Proc. of the AAAI Conf. on Artificial Intell. (AAAI), vol. 35, no.5, pp. 3950–3957, 2021.

[43] R. Hart, L. Yu, Y. Lou and F. Chen, "Improvements on Uncertainty Quantification for Node Classification *via* Distance-based Regularization," NeurIPS, vol. 36, pp. 55454–55478, 2023.

[44] J. Jeong et al., "iGraphMix: Input Graph Mixup Method for Node Classification," Proc. of the 12th Int. Conf. on Learning Representations (ICLR), DOI: 10.1145/3442381.3449796, 2024.

[45] H. Zhao, A. Chen, X. Sun, H. Cheng and J. Li, "All in One and One for All: A Simple Yet Effective Method towards Cross-domain Graph Pre-training," Proc. of the 30th ACM SIGKDD Conf. on Knowledge Discovery and Data Mining (KDD), pp. 4443–4454, DOI: 10.1145/3637528.3671913, 2024.

[46] J. Tang, J. Li, Z. Gao and J. Li, "Re-thinking Graph Neural Networks for Anomaly Detection," Proc. of the 39th Int. Conf. on Machine Learning (ICML), vol.162, pp. 21076–21089, Baltimore, USA, 2022.

[47] M.-H. Guo et al., "PCT: Point Cloud Transformer," Computational Visual Media, vol. 7, pp. 187–199, 2021.

[48] Z. Liu et al., "GraphPrompt: Unifying Pre-training and Downstream Tasks for Graph Neural Networks," Proc. of the ACM on Web Conf. (WWW'23), pp. 417—428, DOI: 10.1145/3543507.3583386, 2023.

[49] X. Yu, C. Zhou, Y. Fang and X. Zhang, "Text-free Multi-domain Graph Pre-training: Toward Graph Foundation Models," [Online], Available: https://doi.org/10.48550/arXiv.2405.13934, 2024.

[50] S. Wang et al., "Multi-domain Graph Foundation Models: Robust Knowledge Transfer *via* Topology Alignment," [Online], Available: https://doi.org/10.48550/arXiv.2502.02017, 2025.

[51] X. Yu et al., "SAMGPT: Text-free Graph Foundation Model for Multi-domain Pre-training and Cross-domain Adaptation," Proc. of the ACM on Web Conf. (WWW), pp. 1142–1153, DOI: 10.1145/3696410.3714828 2025.

[52] Y. Huang et al., "One Prompt Fits All: Universal Graph Adaptation for Pre-trained Models," [Online], Available: https://doi.org/10.48550/arXiv.2509.22416, 2025.

ملخص البحث:

تُعالج هذه الورقة مشكلة النّقل السّلبي في التّدريب المسبق للرّسوم البيانية عبر المجالات في ظلّ سيناريوهات التّعلُّم بعددٍ قليلٍ من الأمثلة، وتقترح إطار عملٍ للتّدريب المسبق متعدّد المكوّنات يُسمّى (نظام تنسيق التّدريب المسبق المعزَّز بالانتباه الخارجي للرسوم البيانية). ويدمج هذا الأطار الانتباه الخارجي متعدّد الرّؤوس مع منسّق الرُّسوم البيانية، علماً بأنّ الأساليب التّقليدية تفتقر للقُدرة على التّكيف مع التّفاعلات المعقّدة والديناميكية، وأنّ معالجة التّباينات الهيكلية والدّلالية بين الرّسوم البيانية عبر المجالات تُعدّ أمراً بالغ الأهمية.

وقد بينت التّجارب على عشر مجموعات بياناتٍ للرّسوم البيانية تنتمي للعالم الحقيقي أنّ النّموذج المقترح أظهر قُدرةً فائقةً على التّعميم عبر المجالات، وأنّه كان ذا أداءٍ مُحَسَّن في مهامّ التّصنيف المتعلّقة بالتّعلُّم بعددٍ قليلً من الأمثلة، مع أفضليةٍ للنّموذج المقترح بلغت 13.3% مقارنةً بالطُّرق الأخرى. حيث يمكن للنّظام المقترح التركّيز بصورةٍ أدقّ على البِنى الحرجة للرّسوم البيانية، موفِّراً بذلك أساساً نظرياً وعملياً لتوظيف الشّبكات العصبية للرّسوم البيانية في السّيناريوهات المعقّدة.

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 12, No. 01, March 2026.

VEGA-360: VIEWPORT-AWARE HIERARCHICAL GROUPED ALLOCATION FOR MULTI-LAYER 360° VIDEO STREAMING

Phi Dinh Huynh and Nguyen Viet Hung

(Received: 26-Nov.-2025, Revised: 13-Jan.-2026, Accepted: 16-Jan.-2026)

ABSTRACT

Delivering multi-layer, tiled 360° video to multiple wireless users is challenging due to limited radio resources, heterogeneous channel conditions and a strong viewport-dependent quality of experience (QoE). To address this problem, we propose Viewport-Enhanced Grouped Allocation for 360° Video (VEGA-360), a hierarchical viewport-aware resource allocation framework for multi-user 360° video streaming. VEGA-360 adopts a two-stage design. On the main stage, VEGA-360 partitions users into a small number of clusters using a joint criterion that combines spectral efficiency and viewport similarity derived from viewport weights and allocates a per-cluster resource budget accordingly. In the fine-tuning phase, VEGA-360 solves independent optimization sub-problems per cluster at the tile granularity, enforcing radio resource and SHVC scalability constraints while maximizing a utility metric that accounts for viewport-weighted visual quality and the transmission overhead caused by distributing multiple tile instances. By separating coarse-grained grouping and budgeting decisions from fine-grained tile-layer allocation, VEGA-360 reduces the size of each optimization instance and improves computational tractability while maintaining viewport-aware service in dense multi-user scenarios. Simulation results show that VEGA-360 achieves competitive utility/QoE compared to a monolithic MILP baseline, with substantially shorter solution times.

KEYWORDS

VEGA-360, QoE, 360-degree video, SHVC encoding, Virtual reality.

1. INTRODUCTION

Virtual Reality (VR) and 360 -degree video technologies have enabled users to experience highly immersive three-dimensional (3D) environments, where they can freely explore and interact with virtual or captured real-world scenes as if they were physically present [1]. By combining head-mounted displays (HMDs), motion tracking and interactive content, VR has been successfully adopted in education, healthcare, manufacturing, entertainment and many other domains, where immersive visual experiences can enhance engagement, training effectiveness and decision-making.

To deliver such immersive experiences at scale, 360 -degree videos are typically captured and rendered in high resolutions (e.g. 4 K and beyond), which leads to extremely large data volumes and high bitrate demands [2]. These requirements become even more stringent on mobile devices, where limited computation capability, constrained battery and fluctuating wireless bandwidth can easily cause playback stalls, blurry viewports or noticeable latency. To cope with these constraints, tile-based viewport-adaptive streaming has emerged as a widely adopted approach: the spherical video is partitioned into multiple tiles and only the tiles within the user's current Field of View (FoV) or Region of Interest (RoI) are delivered at high quality, while the remaining tiles are sent at reduced quality [2]-[3].

Scalable video-coding extensions, such as Scalable High-efficiency Video Coding (SHVC), further enable flexibility by encoding each tile into a base layer and multiple enhancement layers [4]-[5]. The base layer guarantees minimum decodable quality, whereas enhancement layers can be selectively transmitted to refine spatial resolution or improve quality when network conditions permit. At the same time, the evolution of mobile networks from 4G LTE to 5G brings significantly higher peak data rates, lower latency and improved spectral efficiency, offering an attractive infrastructure for delivering interactive 360 -degree video in real time [6]-[7]. As illustrated in Fig. 1, each selected tile is encoded into multiple quality layers and broadcast to different user clusters. High-capability users subscribe to

---

P. D. Huynh and N. V. Hung (Corresponding author) are with the East Asia University of Technology, Vietnam. Emails: 20222072@eaut.edu.vn and hungnv@eaut.edu.vn

more enhancement layers, while low-capability users only receive the base layer. This layered multi-cast structure serves as the basis for the proposed VEGA-360 optimization framework.

Despite these advances, providing high-quality 360 -degree video streaming to multiple mobile users remains challenging. Existing solutions often optimize either viewport-adaptive tiling [2]-[3],[8] or streaming strategies (uni-cast/multi-cast, rate adaptation) [9][10][11][12], but still face difficulties in simultaneously handling: Firstly, heterogeneous link conditions across users; secondly, diverse viewport dynamics and RoI preferences; and finally, limited radio resources on a single cell. Recent studies have explored RoI-based viewport prediction [13]-[14], clustering users with similar viewing patterns [15] and optimized tile-quality selection for multi-user scenarios [16]. However, there is still a lack of a unified framework that jointly exploits user clustering, scalable tiling and resource allocation to maximize Quality of Experience (QoE) under realistic bandwidth constraints.

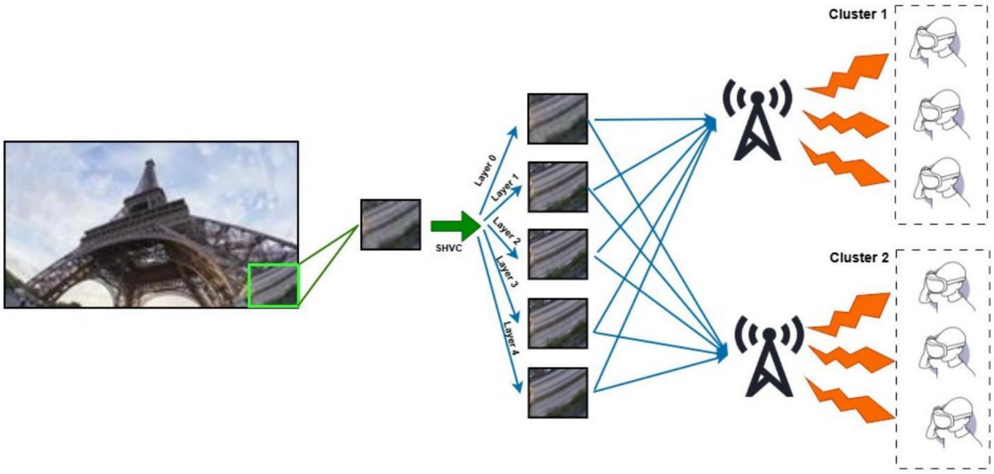


Figure 1. SHVC-based layered multi-cast for tiled 360° video.

Motivated by these limitations, we present VEGA-360, a clustering-based optimization framework for multi-user 360° video streaming over mobile networks. VEGA-360 groups users according to their tile-level quality requirements and channel conditions and allocates radio resources across clusters and tile layers to prioritize high-impact tiles within users' regions of interest (RoIs) while respecting the global bandwidth budget. By jointly leveraging scalable video encoding and intelligent user clustering, VEGA360 effectively balances fairness and efficiency, reduces redundant transmissions and improves overall QoE compared with existing baselines.

The following part provides an expanded overview of the key contributions presented in this paper, highlighting the main ideas, methodological advances and practical implications derived from the study.

- We design a clustering-based system model for multi-user 360 -degree video streaming, where users sharing similar viewport and quality demands are grouped into clusters and tiles are encoded using scalable video coding to support flexible per-tile quality selection.
- We formulate a joint optimization problem that captures tile-layer selection and radio-resource allocation across clusters under bandwidth and QoE constraints, explicitly focusing on tiles lying in users' RoI.
- We develop a practical-solution algorithm to solve the optimization problem segment-by-segment, enabling the system to adapt to time-varying network conditions and dynamic viewports while maintaining stable QoE.
- We conduct extensive simulations using real 360 -degree video traces and head-movement datasets and compare VEGA-360 against other state-of-the-art schemes. The results show that our framework improves QoE and viewport quality while keeping bandwidth usage within practical limits.

This is how the rest of the paper is structured. In Section 2, relevant research on user-clustering techniques and 360 -degree video streaming is reviewed. The system model and problem formulation are presented in Section 3. The suggested framework and solution algorithm are presented in Section 4. Section 5 presents the results of the performance evaluation and Section 6 wraps up the work.

2. RELATED WORK

The rapid growth of immersive services has driven extensive research on adaptive 360 -degree video streaming over wireless networks. Recent work has focused on designing smarter adaptation strategies that consider both network dynamics and user experience. Chen et al. studied streaming 360° VR video with statistical QoS provisioning in mmWave networks, highlighting the role of wireless reliability constraints in immersive delivery [17]. Badnava et al. formulated multi-user 360 -degree video delivery as a multi-task decision-making problem and used a deep reinforcement-learning agent to jointly allocate bitrate and computation resources, with the goal of maximizing long-term QoE under fluctuating bandwidth [18]. Wang et al. adopted a multi-agent deep reinforcement learning framework to control rate adaptation for 360-degree contents, where multiple viewpoints and fairness among users are explicitly modelled in the optimization process [19]. In parallel, Nguyen et al. addressed robustness to sudden throughput reductions and proposed a scalable and resilient 360-degree HTTP/2 streaming solution that exploits stream prioritization and independence to mitigate stalls and quality drops [20]. Other recent studies explored joint optimization of streaming and enhancement. For example, Guo et al. investigated coordinated control of coding parameters and super-resolution filters for mobile 360 -degree delivery [21], while Feng et al. designed a stochastic multi-window adaptation scheme that couples viewport prediction and bitrate assignment [22]. These approaches, however, operate mostly on a per-user basis and do not fully exploit the potential of multi-cast gains when users share similar viewports and quality requirements. FoV overlap has also been explicitly exploited in wireless VR delivery in order to improve robustness and efficiency when users share similar viewing regions [23]. Relatedly, Abedini and Nickray employed reinforcement learning to tune transport-layer congestion control for real-time delivery, indicating that cross-layer adaptation can help stabilize end-to-end performance under time-varying bandwidth [24].

Viewport prediction and user-behavior modeling form another active line of research for 360 -degree video. Wahba et al. provided a recent survey of learning-based viewport-prediction techniques and identified open problems related to latency, generalization and device heterogeneity in practical systems [25]. Building on data-driven prediction, Wang et al. introduced an edge-assisted clustered-learning framework, CoLive, that groups users based on their viewing behaviour and trains cluster-specific models to improve prediction accuracy and streaming efficiency in live scenarios [26]. Zhang et al. proposed a mobile-friendly viewport prediction method for live 360 -degree streaming in which attention-aware features and device constraints are jointly exploited [27]. Besides, Nguyen et al. developed a GRU-LSTM-based viewport-estimation method tailored to 360 -degree video streaming [28]-[29] and more recent work explores reinforcement-learning based viewport estimation that fuses head and eye-movement information (HEVERL) for VR applications [30]. These works clearly show that exploiting correlations between users and leveraging clustering at the prediction layer can improve performance, but they stop short of incorporating clustering directly into a global multi-cast resource-allocation problem.

User clustering for multi-user 360° streaming can be broadly categorized into channel-based and behavior-based clustering. Channel-based clustering groups users according to wireless channel conditions (e.g. spectral efficiency or SNR), which is attractive for multi-cast, because the cluster transmission rate is typically constrained by the worst-channel user. However, clustering solely by channel may ignore viewport heterogeneity and enlarge the union of requested tiles, reducing multi-cast efficiency. In contrast, behavior-based clustering groups users by viewing behaviour (viewport/FoV similarity or predicted viewport trajectories) to maximize tile reuse and reduce redundant transmissions. Representative behavior-aware works emphasized viewport prediction and viewport-adaptive delivery, including SPA360 [31], Meta360 [32], FoV prediction-assisted viewport delivery [33] and utility-driven optimization in JUST360 [34]. While these studies highlighted the benefits of behavior awareness, they generally did not integrate channel heterogeneity into the clustering decision for multi-cast resource allocation. Our work bridges this gap by explicitly combining channel information and viewport similarity in the clustering stage and coupling them with hierarchical per-cluster resource allocation for multi-cast tiled SHVC delivery.

Accurate QoE modeling for immersive media has also received increased attention. More recently, physiological-signal-driven QoE optimization has been investigated for wireless VR transmission, providing an alternative direction for modeling user-perceived experience beyond traditional quality metrics [35]. Nguyen et al. proposed a retina-inspired objective quality-assessment model for tile-coded 360 -degree videos, in which spatial weights are aligned with the non-uniform sensitivity of the human visual system across the field of view [36]. Elwardy et al. presented a pilot study on the consistency of subjective quality assessment for 360 -degree contents, introducing the RQA360 dataset and analyzing repeated tests in both standing and seated viewing conditions [37]. Complementary evidence is reported by Qananwah et al., who explored physiological cues (EEG signals) to guide video-compression decisions, reinforcing the value of human-centric information when constructing QoE-aware streaming and coding strategies [38]. These studies highlighted the importance of QoE metrics that not only reflect the delivered quality level, but also account for viewport importance, spatial-quality variations across tiles and the impact of experimental settings such as viewing posture and device. Such insights motivate the use of viewport-weighted utility functions and penalty terms in the design of optimization-based streaming frameworks. Closer to the present work, clustering-based optimization for 360-degree multi-cast has been investigated from a cross-layer perspective. Nguyen et al. proposed a clustering-based framework for scalable multi-cast of tiled 360-degree videos in multi-cell wireless networks, in which a mixed-integer linear program jointly selects SHVC layers and user clusters under resource constraints [39]. That monolithic approach demonstrates notable QoE and bandwidth gains compared with uni-cast and heuristic baselines, but its computational complexity grows rapidly with the number of users, tiles and available resource blocks, which limits scalability in dense deployments. In contrast, the VEGA-360 framework studied in this paper adopts a hierarchical two-stage design: a first-stage viewport- and channel-aware clustering step groups users and allocates a per-cluster resource budget based on users' spectral efficiencies and viewport similarity (derived from tile weights) and independent second-stage sub-problems allocate tile versions and radio resources within each cluster.

3. SYSTEM MODEL AND DESIGN OVERVIEW

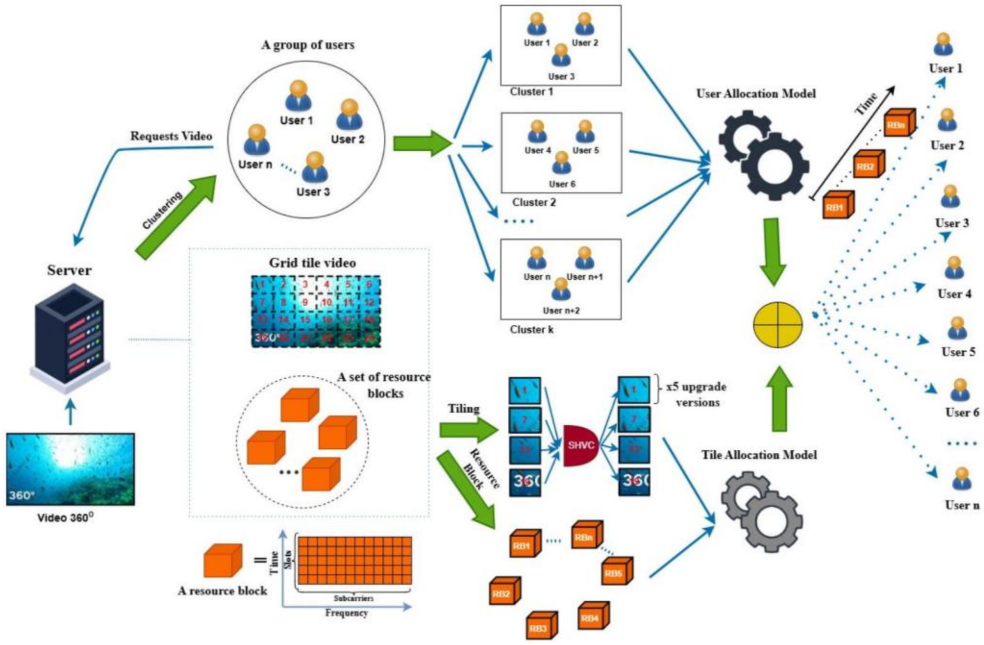


Figure 2. Overall architecture of the proposed VEGA-360 system. The server performs tiling and SHVC encoding, the base station groups users into clusters according to channel quality and viewport similarity and a joint user/tile allocation model maps SHVC layers to resource blocks.

The system model of the suggested VEGA-360 framework for QoE-aware 360-degree video multi-cast, as seen in Fig. 2 and Fig. 3, is described in this section.

We consider a wireless downlink scenario in which a video server stores several 360° videos and communicates with a group of heterogeneous users through a base station (BS). All users request the same 360° video segment at a given time. The processing begins by dividing the video into spatial tiles and temporal segments. Let

$$\mathcal{U} = \{1, \dots, U\}, \mathcal{T} = \{1, \dots, T\} \tag{1}$$

represent the sets of users and tiles within a single segment, respectively. Each tile is encoded by SHVC into C scalable layers, indexed by $c \in \{0, \dots, C-1\}$, where $c = 0$ is the base layer and $c \geq 1$ are enhancement layers. The c-th layer provides an objective video quality Q_c (e.g. PSNR) and requires bitrate (or bandwidth) B_c.

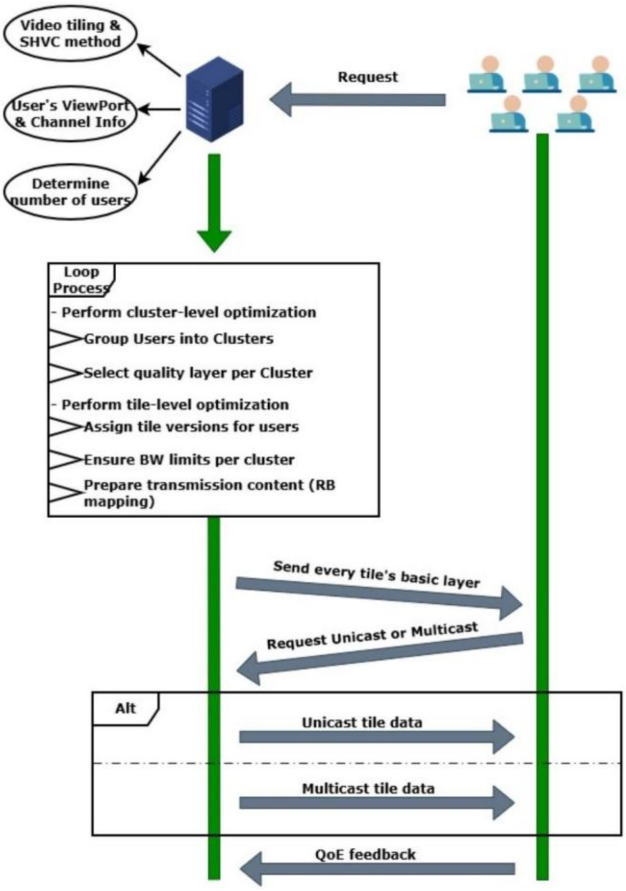

Video tiling &
SHVC method
User's ViewPort
& Channel Info
Determine
number of users
Request
Loop
Process
- Perform cluster-level optimization
Group Users into Clusters
Select quality layer per Cluster
- Perform tile-level optimization
Assign tile versions for users
Ensure BW limits per cluster
Prepare transmission content (RB mapping)
Send every tile's basic layer
Request Unicast or Multicast
Alt
Unicast tile data
Multicast tile data
QoE feedback

Figure 3. Sequence diagram of the VEGA-360 streaming workflow between server and client sides for one video segment.

For each user $u \in \mathcal{U}$, we model the importance of tile $t \in \mathcal{T}$ by a weight $w_{u,t} \in [0,1]$, which reflects how likely tile t lies inside the viewport of user u. The weights are normalized as:

$$\sum_{t \in \mathcal{T}} w_{u,t} = 1, \forall u \in \mathcal{U}. \tag{2}$$

Here, large $w_{u,t}$ means that tile t contributes more to the perceived QoE of user u. The wireless channel of user u is characterized by an average spectral efficiency σ_u (bit/s/Hz), which captures path loss, fading and the selected modulation and coding scheme.

The BS is allocated a total downlink resource budget R per segment, measured, for instance, in OFDM resource blocks (RBs). As depicted in Fig. 2, each RB occupies a certain time-frequency area and can carry coded bits of one or several tile layers. The overall resource constraint is expressed as:

$$\sum_{k=1}^{K} \sum_{c=0}^{C-1} R_{k,c} \leq R, \tag{3}$$

where K is the number of user clusters and $R_{k,c}$ denotes the number of RBs used to transmit the c-th layer that is multi-cast to cluster k. This constraint couples the decisions of quality selection and user grouping across all clusters.

To exploit multi-cast gain while preserving individual QoE, VEGA-360 partitions users into K clusters based on a joint criterion that combines spectral efficiency and viewport similarity derived from $\{w_{u,t}\}$.

$$\mathcal{K} = \{1, \dots, K\}, \tag{4}$$

Users in the same cluster are served by a common multi-cast stream and thus tend to receive similar quality layers, whereas different clusters may receive different numbers of SHVC layers depending on their channel conditions and the global budget R. Clustering reduces the signaling overhead and makes the subsequent optimization scalable when the number of users grows.

The end-to-end operation for one video segment is summarized in Fig. 3. First, the users send a request for a $360°$ video and the server performs tiling and SHVC encoding of the current segment. Second, the BS collects viewport statistics (to update the weights $w_{u,t}$) and channel information (to update the spectral efficiencies σ_u) and determines the current number of active users U. Based on this information, the BS groups users into clusters and allocates a per-cluster resource budget R_k under the total budget R. Next, for every cluster, the BS refines the decision at tile level by assigning, for each tile t and user u in that cluster, which tile version (i.e., which layer index c) should be transmitted, so that all users decode at least the base layer and the per-cluster bandwidth limits are satisfied. Afterwards, the BS prepares the actual transmission content by mapping the selected tile layers onto RBs and delivers them over the air interface. When the same coded tile layer is requested by multiple users in a cluster, it is sent *via* multi-cast; otherwise, individual tiles can be complemented *via* uni-cast when necessary. Finally, after playback, the system may collect QoE-related feedback from users, which can be exploited for long-term adaptation of clustering and resource allocation.

Algorithm 1 summarizes the main-stage clustering of VEGA-360. Unlike purely channel-based grouping, VEGA-360 jointly considers users' spectral efficiencies and viewport similarity derived from the tile-weight vectors $\mathrm{w}_u = \left[w_{u,0}, \dots, w_{u,T-1}\right]$. This design prevents grouping users who have similar channel conditions, but request disjoint viewports, which would otherwise reduce multi-cast gain and waste bandwidth due to union-tile transmissions. After forming clusters, VEGA-360 allocates a per-cluster budget R_k and then performs tile-level optimization within each cluster.

Algorithm 1 Viewport and channel-aware clustering in VEGA-360

```
Collect users' viewport weights $\{w_{u,t}\}$ and spectral efficiencies $\{\sigma_u\}$.
Initialize K clusters $\{U_k\}_{k=0}^{k-1}\}$ with capacities $\{L_k\}$.
Initialize each cluster's centroid viewport vector $\overline{w_k}$ and average channel $\overline{\sigma_k}$ (e.g., using seed users).
for each user u (in descending order of $\sigma_u$ or in any fixed order) do
    for each cluster k with $|U_k| < L_k$ do
        Compute viewport similarity $s_{u,k}^{vp} \leftarrow \mathrm{sim}(w_u, \overline{w_k})$.
        Compute channel similarity $s_{u,k}^{ch} \leftarrow |\sigma_u - \overline{\sigma_k}|$.
        Compute joint score $S_{u,k} \leftarrow \alpha s_{u,k}^{ch} + (1-\alpha)\, s_{u,k}^{vp}$ .
    end for
    Assign user u to the cluster $k^* = \mathrm{argmax}_k S_{u,k}$.
    Update $\overline{w_{k^*}}$ and $\overline{\sigma_{k^*}}$.
end for
Allocate per-cluster budgets $\{R_k\}$ under the total budget R.
return $\{U_k\}$ and $\{R_k\}$.
```

4. PROPOSED VEGA-360 METHOD

4.1 Overview of Tile Transportation

In this sub-section, we present the proposed VEGA-360 framework for QoE-driven multi-cast and uni-cast delivery of tiled $360°$ video. We first use Fig. 4 to explain how VEGA- 360 jointly handles SHVC layers, tiles and time segments. Then, we formulate the main optimization problem and highlight the key constraints that govern user clustering, layer selection and tile-level allocation.

Fig. 4 illustrates the scheduling structure of VEGA-360. The original $360°$ frame is divided into a grid of tiles, indexed by $t \in \mathcal{T} = \{1, \dots, T\}$. Each tile is encoded into C SHVC layers, denoted by $\{v_0, \dots, v_{C-1}\}$, where v_0 is the base layer and higher indices correspond to higher visual quality and higher bitrate.

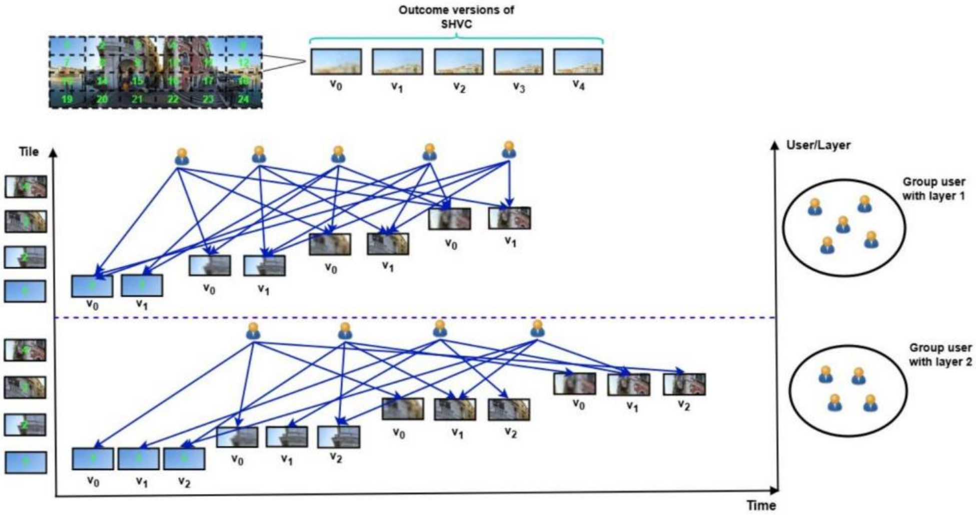


Figure 4. The VEGA-360 structure illustrated with user clusters on the right, time segments on the horizontal and tiles on the vertical.

On the user side, we denote the user set by $\mathcal{U} = \{1, \ldots, U\}$. For each user $u \in \mathcal{U}$ and each tile $t \in \mathcal{T}$, we pre-compute a weight $w_{u,t}$ that reflects how likely tile t falls into the user's viewport. Tiles that frequently appear in the viewport are assigned larger weights, while tiles rarely watched by user u have very small weights. Along the horizontal axis in Fig. 4, time is divided into segments (chunks) of equal duration. For every segment, VEGA-360 first groups users into clusters and allocates a per-cluster resource budget R_k and then decides which SHVC layer is transmitted for each user-tile pair within each cluster.

Users with similar channel conditions and viewport characteristics are dynamically grouped into clusters. In Fig. 4, the right-hand side shows two such clusters. Inside each cluster, the base layer v_0 of the required tiles is always transmitted to guarantee decodability, while enhancement layers $v_1, \ldots, v_{C-1}$ are selectively delivered depending on viewport weights and channel conditions, as depicted by the blue arrows.

4.2 Main Formulae Contribution

For each user $u \in \mathcal{U}_k$, tile $t \in \{0, \ldots, T-1\}$ and version index $c \in \{0, \ldots, C_k - 1\}$, we introduce a binary decision variable

$$y_{u,t,c} = \begin{cases} 1, \text{ if version } c \text{ of tile } t \text{ is delivered to user } u, \\ 0, \text{ otherwise}. \end{cases} \tag{5}$$

The per-cluster viewport quality for one segment is computed as:

$$VQ_k = \sum_{u \in \mathcal{U}_k} \sum_{t=0}^{T-1} \sum_{c=0}^{C_k - 1} w_{u,t} Q_c y_{u,t,c}, \tag{6}$$

where VQ_k measures the accumulated quality perceived by all users in cluster $k, w_{u,t}$ is the normalized importance of tile t in the viewport of user u and Q_c denotes the objective quality (e.g. PSNR) of version c.

In parallel, we keep track of how many tile versions are transmitted in cluster :

$$TV_k = \sum_{u \in \mathcal{U}_k} \sum_{t=0}^{T-1} \sum_{c=0}^{C_k - 1} y_{u,t,c}, \tag{7}$$

where TV_k acts as a proxy for transmission overhead and decoding complexity, since each selected version corresponds to an additional bitstream that must be sent and decoded.

The global QoE metric of VEGA-360 combines viewport quality and transmission cost as:

$$d_k \triangleq \frac{\tau_{\text{seg}}}{R_k} \sum_{u \in U_k} \sum_{t} \sum_{c} \frac{\left(y_{u,t,c} - z_{u,t,c}\right) B_c}{\sigma_u} \tag{8}$$

$$r_u \geq d_{k(u)} - b_u, r_u \geq 0 \tag{9}$$

$$\text{QoE} = \frac{\alpha}{1000} \sum_{k=0}^{K-1} VQ_k - \frac{\gamma}{1000} \sum_{k=0}^{K-1} TV_k - \frac{\beta}{1000} \sum_{u=1}^{U} r_u \tag{10}$$

where $\alpha > 0$ weights viewport quality and $\gamma > 0$ penalizes transmission cost (number of delivered tile versions). d_k is the effective delivery time of cluster k for one segment and r_u denotes the rebuffering time of user u, modelled by the linear constraints in Eq. (9) (i.e., $r_u = \left[d_{k(u)} - b_u\right]^+$). $\beta > 0$ weights the stalling penalty. We set $\alpha = 1, \gamma = 1$ and $\beta = 1.85$ and apply a normalization factor of $1/1000$ for numerical stability.

Each cluster is subject to its own radio-budget constraint. The resource consumption of cluster k is upper bounded by R_k:

$$\sum_{u \in \mathcal{U}_k} \sum_{t=0}^{T-1} \sum_{c=0}^{C_k-1} \frac{B_c}{\sigma_u} y_{u,t,c} \leq R_k, \tag{11}$$

where B_c is the bitrate of version c, σ_u denotes the spectral efficiency of user u and R_k is the share of available radio resources allocated to cluster k. This constraint guarantees that the cumulative bandwidth of the selected tile versions does not surpass the cluster's allocated budget.

In our implementation, the per-cluster budget R_k is obtained by splitting the total budget R proportionally to the minimum base-layer delivery cost of each cluster. Specifically, we compute

$$W_k = \sum_{u \in U_k} \frac{B_0}{\sigma_u}, R_k = R \cdot \frac{W_k}{\sum_{j=0}^{K-1} W_j}, \tag{12}$$

where B_0 is the bitrate of the base layer. This allocation assigns more resources to clusters with poorer channel conditions (smaller σ_u), ensuring base-layer feasibility and avoiding starvation of low-capability users before the tile-level optimization in Eq. (16).

Due to the hierarchical structure of SHVC, an enhancement version can only be decoded if all lower versions of the same tile are also available. VEGA-360 enforces this dependency through

$$y_{u,t,c} \leq y_{u,t,c-1}, \forall u \in \mathcal{U}_k, \forall t, c = 1, \ldots, C_k - 1, \tag{13}$$

which ensures that whenever version c is selected, version $c - 1$ is selected as well. Moreover, the basic visibility of the $360°$ scene is always guaranteed by forcing the base layer of every tile to be sent:

$$y_{u,t,0} = 1, \forall u \in \mathcal{U}_k, \forall t. \tag{14}$$

Finally, all decision variables are binary,

$$y_{u,t,c} \in \{0,1\}, \forall u \in \mathcal{U}_k, \forall t, \forall c, \tag{15}$$

so that each version of each tile is either fully selected or not transmitted. Putting everything together, the tile-level optimization in cluster k is written as

$$\max_{\{y_{u,t,c}\}} \alpha VQ_k - \gamma TV_k \text{ s.t. Eq.11-15,} \tag{16}$$

and is solved for every cluster under its corresponding budget R_k. To further exploit the heterogeneity of viewports inside a cluster, VEGA-360 refines the solution of Eq. (16) by imposing a QoE-aware ordering across users, as summarized in Algorithm 2. For each tile t, the users in $\mathcal{U}_k$ are first sorted in descending order of their viewport weights $w_{u,t}$, obtaining a sequence $\left(u_0, u_1, \ldots, u_{|\mathcal{U}_k|-1}\right)$ from the most to the least-interested user. Then, for every adjacent pair (u_i, u_{i+1}) and every admissible version c, the inequality

$$y_{u_{i+1},t,c} - y_{u_i,t,c} \leq 0$$

is enforced. This simple rule guarantees that a user with lower importance for tile t never receives a higher version than a user with higher importance in the same cluster. As a result, VEGA-360 creates a staircase pattern of tile versions inside each cluster: central viewports are upgraded first, while the multicast structure is preserved, leading to a better QoE-bandwidth trade-off.

Algorithm 2 QoE-aware ordering of tile versions in cluster k

```
for t = 0 to T − 1 do
    Sort users in $u_k$ by descending $w_{u,t}$ obtain the sequence $(u_1, u_2, . . . , u_{|u_k|-1})$
    for i = 0 to $|u_k| - 1$ do
        if i + 1 < $|u_k|$ then
            for c = 0 to $C_k - 1$ do
                $y_{u_{i+1},t,c} - y_{u_i,t,c} \leq 0$
            end for
        end if
    end for
end for
```

5. PERFORMANCE EVALUATION

In this section, we first describe the 360 -degree video dataset, encoding configuration and simulation parameters used to evaluate VEGA-360 and then discuss the obtained performance in comparison with existing baselines.

5.1 Experimental Setup

Our performance evaluation is carried out in a custom simulator implemented in Python, where we replay the clustering and scheduling decisions of VEGA-360 together with the baseline algorithms. All mixed integer programs are solved by the Gurobi optimizer on a standard workstation equipped with an Intel Core i7 CPU and 32 GB of RAM, with a moderate time limit per instance.

We reuse a public 360 -degree video dataset introduced in [40], which contains five omni-directional sequences: Rollercoaster, Diving, Venice, Paris and Rhino. Following the original dataset, the videos are grouped into two content categories: "less-feature" (mostly static scenes) and "more-feature" (dynamic scenes with many moving objects and camera motion). Each raw video is projected to the equirectangular format with a resolution of 2890×1920 pixels and partitioned into $T = 24$ tiles of size 480×480 pixels per tile. To support scalable streaming, every tile is encoded with the SHVC extension of HEVC into one base layer and four enhancement layers, i.e., $C = 5$ quality versions per tile. Table 1 summarizes the average viewport PSNR (in dB) and bitrate (in kbps) of the five versions for the two content categories, averaged over all videos in the dataset.

We consider a single cell serving $U = 70$ mobile users requesting the same 360 -degree video. User-specific viewport trajectories are obtained from the head-movement traces. For each segment, we project the viewport field-of-view onto the tiled equirectangular plane and compute tile weights by the normalized viewport-tile overlap ratios, so that the weights reflect how much each user watches each tile in that segment. Following the COSMN setting, we generate $U = 70$ users by sampling traces with replacement and randomizing the starting segment index to avoid synchronized viewing patterns. The wireless downlink is abstracted by a total resource budget per segment shared by all users and clusters and the budget is swept from 10000 to 120000 (arbitrary resource units) to emulate different congestion levels. Users' spectral efficiencies follow the same channel abstraction as COSMN [39] and are used to determine per-user transmission cost in the optimization. For the rebuffering term, we use a fixed segment duration and a fixed initial playback buffer as Eq. (17):

$$\tau_{\text{seg}} = 1 \text{ s}, b_u = b_0 = 2 \text{ s}, \forall u \in \{1, \dots, U\}. \quad (17)$$

For each value of R, we run VEGA-360 and four baseline schemes under the same $w_{u,t}$ and encoding parameters:

- COSMN [39]: the original clustering-based optimization that solves a single global MILP over all users, tiles and versions.

- LVSUM [9]: a greedy layer-by-layer strategy that upgrades tile versions sequentially according to the remaining budget.
- Multi-cast All [41]: a multi-cast-only scheme that delivers the same version of each tile to every user, without exploiting viewport diversity.
- Multi-cast Sca [10]: a scalable multi-cast scheme that uses the SHVC layers, but still ignores fine-grained viewport heterogeneity.

Table 1. Average viewport PSNR and bitrate of the encoded tile versions.

Values	Ver. 0	Ver. 1	Ver. 2	Ver. 3	Ver. 4
Less-feature videos					
PSNR (dB)	36.79	41.09	42.89	45.64	48.30
Bitrate (kbps)	58.57	148.37	281.69	560.84	995.61
More-feature videos					
PSNR (dB)	35.30	38.70	41.45	44.71	47.34
Bitrate (kbps)	177.40	417.32	741.89	1357.76	2143.49

All schemes are evaluated using the QoE metric defined in Eq. (10), which linearly trades-off viewport quality against the number of transmitted tile versions. In addition, we also report the average viewport PSNR (VQ) and the average number of transmitted tile versions per segment (TV) as auxiliary indicators.

5.2 Results and Discussion

On the one hand, Table 2 displays the average quality of experience measured for the video with fewer features. VEGA-360 achieves consistently strong QoE across feasible bandwidth budgets and it remains competitive (often best) when the bandwidth budget increases. This behavior indicates that VEGA-360 can sustain user-perceived quality even when user population grows, while still respecting the available radio resources. In addition, the performance gap among optimization-based methods becomes small at medium-to-high budgets, suggesting that QoE gradually saturates once most viewport-important tiles can be delivered at adequate quality.

Table 2. Average QoE for the "less-feature" video under different bandwidth budgets and numbers of users.

Method	Users	Bandwidth budget (kRBs)						
		20	30	45	55	65	75	80
VEGA-360	35	-	-	1.949	1.955	1.969	1.966	1.971
	70	-	-	2.610	2.614	2.632	2.625	2.637
COSMN	35	1.924	1.936	1.947	1.954	1.960	1.965	1.967
	70	2.584	2.596	2.607	2.613	2.620	2.624	2.626
LVSUM	35	1.917	1.927	1.943	1.949	1.957	1.962	1.964
	70	2.577	2.586	2.602	2.609	2.616	2.621	2.623
Multicast All	35	-	-	-	-	-	1.873	1.880
	70	-	-	-	-	-	2.532	2.540
Multicast Sca	35	-	-	-	-	-	1.880	1.891
	70	-	-	-	-	-	2.540	2.550

On the other hand, Table 3 summarizes the results for the more-feature video. Compared to the less-feature case, the more-feature content typically requires higher delivery effort to maintain the same perceived quality, which makes the bandwidth budget more influential. VEGA-360 remains robust across a broader range of budgets and user scales, demonstrating stable QoE when scaling from 35 users to 70 users.

Overall, the results highlight that the proposed design maintains competitive QoE under heterogeneous user demands and richer visual details.

To better illustrate the resource aspect, Fig. 5 and Fig. 6 visualize the bandwidth consumption under different user populations. In general, as the number of users increases, purely unicast-oriented strategies tend to incur higher bandwidth usage. By contrast, VEGA-360 is designed to exploit multi-cast opportunities through QoE-aware ordering and reuse of tile versions, thereby limiting redundant transmissions while preserving viewport quality. This multicast-aware behavior is more evident in the more-feature setting, where the content complexity and quality requirements amplify the benefit of coordinated version delivery.

Finally, Fig. 8 and Fig. 7 present the PSNR comparison across bandwidth budgets for each method. Overall, VEGA-360 achieves competitive (and often higher) reconstruction quality, especially in the low-to-medium budget region where bandwidth scarcity makes tile/version prioritization critical. As the budget increases, the PSNR gap among optimization-based schemes becomes smaller, indicating a saturation effect: once most viewport-relevant tiles can be delivered at sufficiently high quality, additional bandwidth yields marginal visual gains. These results support the main goal of VEGA-360, i.e., improving visual fidelity (PSNR) while avoiding unnecessary bandwidth inflation in 360° tiled streaming.

Table 3. Average QoE for the "more-feature" video under different bandwidth budgets and numbers of users.

Method	Users	Bandwidth budget (kRBs)						
		50	80	120	180	250	280	300
VEGA-360	35	-	1.901	1.905	1.922	1.929	1.935	1.932
	70	-	2.516	2.541	2.547	2.570	2.576	2.576
COSMN	35	1.870	1.889	1.902	1.919	1.929	1.931	1.931
	70	2.515	2.533	2.547	2.564	2.573	2.575	2.576
LVSUM	35	1.858	1.879	1.896	1.915	1.926	1.929	1.930
	70	2.503	2.524	2.541	2.560	2.571	2.574	2.575
Multicast All	35	-	-	-	-	1.828	1.835	1.840
	70	-	-	-	-	2.473	2.480	2.484
Multicast Sca	35	-	-	-	-	1.838	1.849	1.855
	70	-	-	-	-	2.483	2.494	2.500

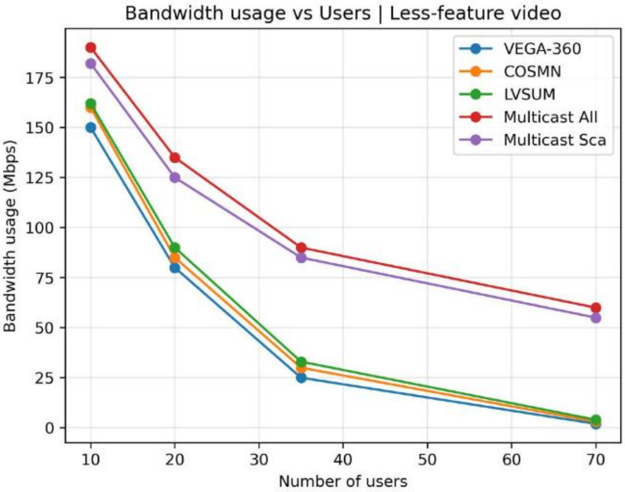


Figure 5. Bandwidth consumption *versus* number of users for the "less-feature" video. VEGA-360 consistently requires lower bandwidth than COSMN and other baselines under the same user load.

We also present Table 4 that reports the average solver runtime of our method and the baselines under different total radio budgets R. Overall, our method consistently achieves lower runtime than COSMN across all tested R values (bold entries). In particular, while COSMN requires about $0.55 - 1.00$ second, our method finishes within $0.28 - 0.55$ second, showing a clear reduction in decision latency as the budget increases. The runtimes of the other baselines (LMSUM, Mul_all and Mul_sca) lie between COSMN and our method in most cases.

After analyzing the outcomes, we identify several practical remarks emerging from the experimental evaluation:

- First, the benefits of VEGA-360 are most evident in the low-to-medium bandwidth region, where resource scarcity forces strict prioritization between tiles and instances; in such regimes, coordinated instance reuse and multi-cast distribution avoid redundant transmissions while maintaining image quality.
- Second, the PSNR curves tend to saturate as bandwidth budgets increase, which suggests diminishing returns once the majority of view-relevant tiles are already delivered at sufficiently high quality; thus, pursuing aggressive upgrades at high budgets is less beneficial than improving efficiency at tight budgets.
- Third, scaling the number of users amplifies the advantages of multi-directional awareness designs: while unidirectional-focused strategies will generate bandwidth growth that is roughly proportional to the number of users, VEGA-360 can limit additional bandwidth by reusing cell instances within groups whenever users share similar viewing preferences.

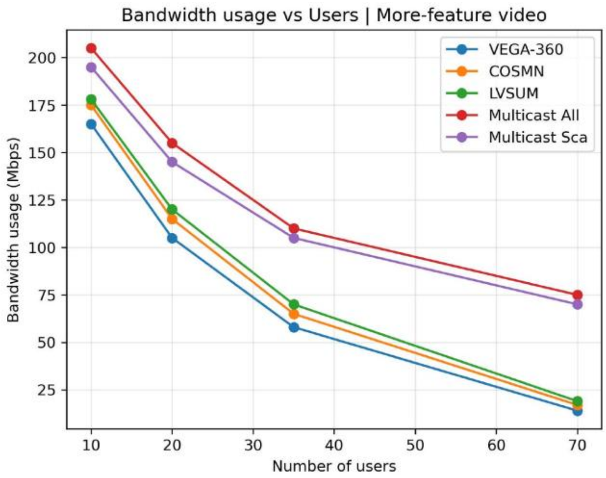


Figure 6. Bandwidth consumption *versus* number of users for the "more-feature" video. The proposed VEGA-360 maintains the best bandwidth efficiency across all evaluated user scales.

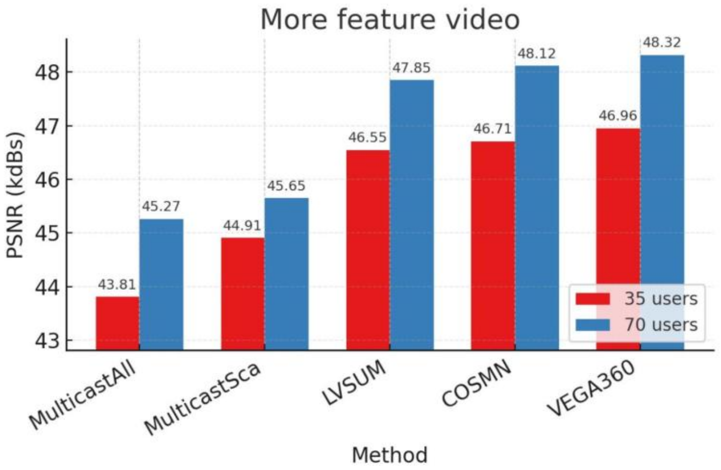


Figure 7. Viewport-quality comparison for the "more-feature" video in terms of PSNR. Results are shown for 35 users and 70 users, highlighting VEGA-360's advantage over COSMN and multi-cast baselines.

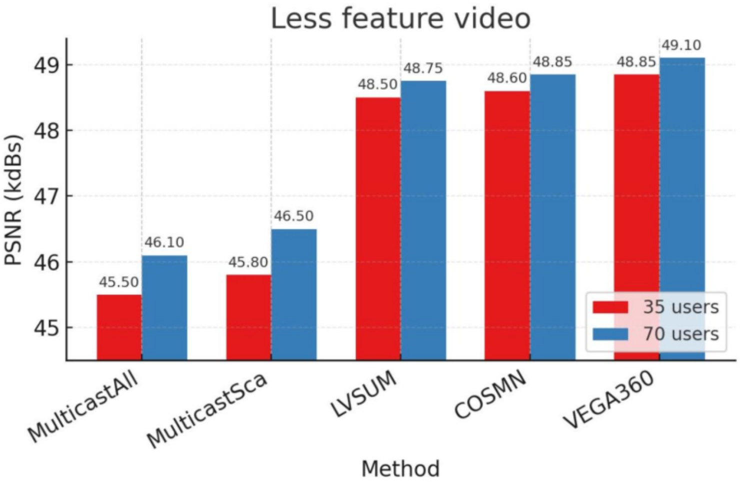


Figure 8. Viewport quality comparison for the "less-feature" video in terms of PSNR. VEGA360 achieves the highest PSNR for both 35- and 70-user scenarios.

- Finally, these observations indicate that VEGA-360 is particularly well-suited to scenarios with multiple users and bandwidth-constrained wireless systems, where improving the quality-efficiency trade-off has a larger impact than marginal quality gains at resfource-rich levels.

Table 4. Runtime comparison (second) under different total radio budgets R.

R	COSMN	LMSUM	Mul_all	Mul_sca	Our method
20000	0.55	0.48	0.30	0.44	0.28
30000	0.62	0.54	0.33	0.48	0.31
45000	0.71	0.61	0.47	0.53	0.36
55000	0.82	0.70	0.51	0.57	0.41
65000	0.90	0.78	0.64	0.68	0.46
75000	0.96	0.83	0.66	0.73	0.52
80000	1.00	0.86	0.78	0.85	0.55

6. CONCLUSION

In this paper, we present VEGA-360, a QoE-aware delivery framework for tile-based 360° video streaming over bandwidth-constrained wireless networks. VEGA-360 is designed to balance image quality and transmission performance by coordinating tile-based version selection and leveraging multi-cast opportunities across user groups. Experimental results on representative 360° content across different bandwidth budgets and user sizes show that VEGA-360 achieves competitive performance compared to baseline schemes, especially in bandwidth-constrained regimes where efficient reuse of delivered versions is crucial. As a two-stage framework, VEGA-360 trades global optimality for computational tractability and the clustering decision may introduce a small performance gap compared to a monolithic formulation in certain cases. In addition, our current setting assumes a shared-content scenario where users request the same 360° video; multi-cast gains may decrease in heterogeneous on-demand scenarios.

In the future, we plan to extend VEGA-360 in three directions: first, integrating online-view prediction and prediction-fault tolerance; next, exploring adaptive clustering and dynamic update intervals under rapidly changing channel and mobile user conditions; and finally, deploying a real-time prototype to evaluate end-to-end delay and system overhead in real networks. We will also investigate more flexible grouping mechanisms (e.g. adaptive or soft cluster-size constraints) to further improve robustness and efficiency.

REFERENCES

[1] A. J. Nair et al., "Unleashing Digital Frontiers: Bridging Realities of Augmented Reality, Virtual Reality and the Metaverse," The Metaverse Dilemma: Challenges and Opportunities for Business and Society, Emerald Publishing Limited, p. 122024, DOI: 10.1108/978-1-83797-524-220241006, 2024.

[2] J. Tu et al., "Pstile: Perception-sensitivity-based 360° Tiled Video Streaming for Industrial Surveillance," IEEE Transactions on Industrial Informatics, vol. 19, no. 9, pp. 9777-9789, 2023.

[3] K. K. Sreedhar et al., "Viewport-adaptive Encoding and Streaming of 360-degree Video for Virtual Reality Applications," Proc. of the 2016 IEEE (ISM), pp. 583-586, San Jose, USA, 2016.

[4] C.-H. Yeh et al., "Fast Prediction for Quality Scalability of High Efficiency Video Coding Scalable Extension," Journal of Visual Communication and Image Representation, vol. 58, pp. 462-476, 2019.

[5] J. M. Boyce et al., "Overview of SHVC: Scalable Extensions of the High Efficiency Video Coding Standard," IEEE Trans. on Circuits and Systems for Video Technology, vol. 26, no. 1, pp. 20-34, 2015.

[6] M. J. Mohammed et al., "A Comparison of 4G LTE and 5G Network Cybersecurity Performance," Proc. of the 2024 35th Conf. of Open Innovations Association (FRUCT), pp. 452-464, Tampere, Finland, 2024.

[7] L. Sun et al., "Multi-path Multi-tier 360-degree Video Streaming in 5G Networks," Proc. of the 9th ACM Multimedia Systems Conf., pp. 162-173, DOI: 10.1145/3204949.3204978, 2018.

[8] N. Al-Najdawi, "High Performance Block Matching Algorithm for High Bit-rate Real-time Video Communication," Jordanian J. of Computers and Inf. Tech. (JJCIT), vol. 4, no. 1, pp. 10-24, 2018.

[9] N. V. Hung et al., "LVSUM-Optimized Live 360 Degree Video Streaming in Unicast and Multicast over Mobile Networks," Proc. of the 2023 IEEE 15th Int. Conf. on Computational Intelligence and Communication Networks (CICN), pp. 29-34, Bangkok, Thailand, 2023.

[10] D. Nguyen, N. V. Hung, N. T. Phong, T. T. Huong and T. C. Thang, "Scalable Multicast for Live 360-degree Video Streaming over Mobile Networks," IEEE Access, vol. 10, pp. 38802-38 812, 2022.

[11] D. T. Nguyen, T. H. Tran and V. H. Nguyen, "Mellifluous Viewport Bitrate Adaptation for 360° Videos Streaming over HTTP/2," ICT Research, [Online], Available: https: //api.semanticscholar.org/CorpusID: 273740657, 2024.

[12] N. V. Hung, B. D. Tien, T. T. T. Anh, P. N. Nam and T. T. Huong, "An Efficient Approach to Terminate 360-video Stream on HTTP/3," Proc. of AIP Conf. Proc., vol. 2909, no. 1, p. 050004, AIP Publish, 2023.

[13] X. Feng, W. Li and S. Wei, "LiveROI: Region of Interest Analysis for Viewport Prediction in Live Mobile Virtual Reality Streaming," Proc. of the 12th ACM Multimedia Systems Conf., pp. 132-145, DOI: 10.1145/3458305.3463378, 2021.

[14] N. Hung et al., "Building an Online Learning Model through a Dance Recognition Video Based on Deep Learning," Informatics and Automation, vol. 23, no. 1, pp. 101-128, 2024.

[15] M. R. Civanlar, Proc. of the 30th ACM Workshop on Network and Operating Systems Support for Digital Audio and Video, ser. ACM Conferences, New York, NY: Association for Computing Machinery, 2020.

[16] M. Mahmoud et al., "Optimized Tile Quality Selection in Multi-user 360° Video Streaming," IEEE Open Journal of the Communications Society, vol. 5, pp. 7301-7316, 2024.

[17] Y. Chen et al., "Streaming 360° VR Video with Statistical QoS Provisioning in mmWave Networks from Delay and Rate Perspectives," IEEE Trans. on Wireless Comm., vol. 24, no. 6, pp. 4721-4737, 2025.

[18] B. Badnava, J. Chakareski and M. Hashemi, "Multi-task Decision-making for Multi-user 360 Video Processing over Wireless Networks," Proc. of the 2024 IEEE 7th Int. Conf. on Multimedia Information Processing and Retrieval (MIPR), pp. 294-300, DOI: 10.1109/MIPR62202.2024.00054, 2024.

[19] H. Wang, Z. Long, H. Dong and A. El Saddik, "MADRL-based Rate Adaptation for 360° Video Streaming with Multi-viewpoint Prediction," IEEE IoT J., vol. 11, no. 15, p. 26503-26517, Aug. 2024.

[20] V. H. Nguyen et al., "Scalable and Resilient 360-degree-video Adaptive Streaming over HTTP/2 against Sudden Network Drops," Computer Communications, vol. 216, pp. 1-15, 2024.

[21] H. Guo et al., "Joint Adaptation for Mobile 360-degree Video Streaming and Enhancement," IEEE Transactions on Mobile Computing, vol. 24, no. 8, pp. 7726-7741, 2025.

[22] W. Feng, S. Wang and Y. Dai, "Adaptive 360-degree Streaming: Optimizing with Multi-window and Stochastic Viewport Prediction," IEEE Trans. on Mobile Computing, vol. 24, no. 7, pp. 5903-5915, 2025.

[23] J. Lee, H. Lu and Y. Chen, "Robust Wireless VR Video Transmission Based on Overlapped FoVs," Proc. of the IEEE Int. Conf. on Communications (ICC 2023), pp. 3084-3089, Rome, Italy, 2023.

[24] M. N. Ehsan Abedini, "Cubic-learn: A Reinforcement Learning Approach to Cubic Congestion Control," Jordanian Journal of Computers and Information Technology (JJCIT), vol. 11, no. 4, pp. 466 483, 2025.

[25] M. Z. A. Wahba, S. Baldoni and F. Battisti, "Learning-based Viewport Prediction for 360-degree Videos: A Review," Electronics, vol. 14, no. 18, p. 3743, 2025.

[26] M. Wang et al., "CoLive: Edge-assisted Clustered Learning Framework for Viewport Prediction in 360° Live Streaming," IEEE Transactions on Multimedia, vol. 26, pp. 5078-5091, 2024.

[27] L. Zhang et al., "Optimizing Mobile-friendly Viewport Prediction for Live 360-degree Video Streaming," IEEE Transactions on Mobile Computing, vol. 24, no. 10, pp. 10441-10455, 2025.

[28] H. Nguyen et al., "An Accurate Viewport Estimation Method for 360 Video Streaming Using Deep Learning," EAI Endorsed Trans. on Industrial Networks and Intelligent Systems, vol. 9, no. 4, p. e2, 2022.

[29] N. V. Hung et al., "Building Predictive Smell Models for Virtual Reality Environments," Computer Science and Automation, vol. 24, no. 2, pp. 556-582, 2025.

[30] N. V. Hung et al., "HEVERL: Head-eye Movement Oriented Viewport Estimation Based on Reinforcement Learning," Journal on VR/Multimedia Systems, 2025, "Viewport Prediction for 360-

degree Video Using Reinforcement Learning," Russian Academy of Sciences, 2025.

[31] Y. Wang et al., "Synergistic Temporal-spatial User-aware Viewport Prediction for Optimal Adaptive 360-degree Video Streaming," IEEE Transactions on Broadcasting, vol. 70, no. 2, pp. 453-467, 2024.

[32] J. Li, Y. Wang and Y. Liu, "Meta360: Exploring User-specific and Robust Viewport Prediction in 360-degree Videos through Bi-directional LSTM and Meta-adaptation," Proc. of the 2023 IEEE Int. Symposium on Mixed and Augmented Reality (ISMAR), pp. 652-661, Sydney, Australia, 2023.

[33] A. Yaqoob and G.-M. Muntean, "A Combined Field-of-view Prediction-assisted Viewport Adaptive Delivery Scheme for 360° Videos," IEEE Trans. on Broadcasting, vol. 67, no. 3, pp. 746-760, 2021.

[34] Z. Li, Y. Wang, Y. Liu, J. Li and P. Zhu, "JUST360: Optimizing 360-degree Video Streaming Systems with Joint Utility," IEEE Transactions on Broadcasting, vol. 70, no. 2, pp. 468-481, 2024.

[35] C. Wu, Y. Chen, Y. Chen, F. Guo, X. Qin and H. Lu, "Physiological Signal-driven QoE Optimization for Wireless Virtual Reality Transmission," arXiv: 2508.09151, DOI: 10.48550/arXiv.2508.09151, 2025.

[36] V. H. Nguyen et al., "Retina-based Quality Assessment of Tile-coded 360-degree Videos," EAI Endorsed Transactions on Industrial Networks and Intelligent Systems, vol. 9, no. 32, p. e2, 2022.

[37] M. Elwardy et al., "On the Consistency of 360 Video Quality Assessment in Repeated Subjective Tests: A Pilot Study," EAI Endorsed Trans. on Industrial Networks and Intell. Systems, vol. 11, no. 1, Jan. 2024.

[38] R. B. Qasem Qananwah et al., "The Utilization of EEG Signal in Video Compression," Jordanian Journal of Computers and Information Technology (JJCIT), vol. 5, no. 3, pp. 263-274, 2019.

[39] H. N. Viet et al., "COSMN: Clustering-based Optimization for 360-degree Live Streaming over Mobile Networks," EAI Endorsed Trans. on Industrial Networks and Intell. Sys., vol. 13, no. 1, pp. 1-16, 2025.

[40] X. Corbillon et al., "360-degree Video Head Movement Dataset," Proc. of the 8th ACM on Multimedia Systems Conf. (MMSys'17), pp. 199-204, DOI: 10.1145/3083187.3083215, 2017.

[41] D. Nguyen, N. V. Hung, T. T. Huong and T. C. Thang, "A Cross-layer Framework for Multi-user 360-degree Video Streaming over Cellular Networks," Proc. of the 2022 IEEE Int. Conf. on Consumer Electronics (ICCE), DOI: 10.1109/ICCE53296.2022.9730536, Las Vegas, USA, 2022.

ملخص البحث:

يُعدّ بثّ الفيديو بزاوية 360 درجة متعدّد الطّبقات إلى عدّة مستخدمين تحدّياً بالنّظر إلى محدودية موارد الرّاديو، واختلاف ظُروف القنوات، وتأثُّر جودة تجربة المستخدم بشكلٍ كبير بحجم شاشة العرض. ولمعالجة هذه المشكلة، نقترح في هذه الدّراسة نظاماً لتخصيص الموارد يُراعي حجم شاشة العرض لبثّ الفيديو إلى عدّة مستخدمين بزاوية 360 درجة. ويعتمد النّظام المقترح تصميماً ثنائي المراحل؛ المرحلى الأولى هي الرئيسية، وفيها يتمّ تقسيم المستخدمين إلى عددٍ من المجموعات ويخصّص النّظام لكلّ مجموعة ميزانية موارد. أمّا المرحلة الثانية، فهي مرحلة الضّبط الدّقيق، وفيها يتمّ تعظيم مقياس يُراعي جودة الصّورة بالإضافة إلى عبء الإرسال، محسِّناً بذلك من سهولة الحساب، مع الحفاظ على خدمةٍ فراغيةٍ لِمَنْفَذِ العرض في سيناريوهات تعدُّد المستخدمين بشكلٍ كثيف.

وتُظهر نتائج المحاكاة أنّ النّظام المقترح يحقّق جودة تجربةٍ تنافسية عالية عند مقارنته بأنظمةٍ أخرى مماثلة واردة في أدبيات الموضوع، مع أوقات حلٍّ أقصر بشكلٍ ملحوظ.

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 12, No. 01, March 2026.

A SCALABLE FEDERATED DEEP REINFORCEMENT LEARNING ARCHITECTURE FOR COLLABORATIVE LEARNING

Tarek Amine Haddad

(Received: 29-Dec.-2025, Revised: 8-Feb.-2026, Accepted: 10-Feb.-2026)

ABSTRACT

Federated Learning enables collaborative model training without sharing raw data, while Deep Reinforcement Learning provides powerful mechanisms for sequential decision-making. However, their integration suffers from limited scalability, sensitivity to non-IID data and unstable convergence in distributed environments. This paper proposes a Scalable Federated Deep Reinforcement Learning (SFDRL) architecture in which distributed agents learn local policies and periodically contribute to a global model via an adaptive, performance-aware aggregation strategy. Unlike conventional FedRL methods that rely on uniform averaging, SFDRL weights local updates according to their learning effectiveness, resulting in faster convergence and improved stability under heterogeneous data distributions. In addition, a selective communication mechanism is introduced to reduce communication overhead by up to 28% and 64% compared with FedAvg and FedRL, respectively. Extensive experiments demonstrate that SFDRL outperforms compared methods, achieving higher cumulative rewards, reduced variance during training and improved scalability in large-scale distributed settings. These results confirm the suitability of SFDRL for practical deployment in distributed intelligent systems.

KEYWORDS

Federated learning, Deep reinforcement learning, Collaborative learning, Distributed intelligence, Scalability, Adaptive aggregation.

1. INTRODUCTION

Deep Reinforcement Learning (DRL) has emerged as a powerful paradigm for sequential decision-making in complex and high-dimensional environments. By integrating reinforcement learning principles with deep neural networks, DRL enables agents to learn optimal control policies directly from raw sensory inputs without relying on handcrafted features. This capability has led to remarkable successes in a wide range of applications, including robotics, autonomous driving, intelligent transportation systems and resource management. Value-based methods, such as Deep Q-Networks (DQNs) and policy-based or actor-critic approaches like DDPG and PPO, have demonstrated strong performance in both discrete and continuous action spaces. Despite these advances, conventional DRL typically relies on centralized training with full access to experience data, which limits its scalability and raises privacy and communication concerns in distributed and multi-agent environments [1]. In addition, DRL has demonstrated remarkable success in various domains, including robotics [2], intelligent transportation [3] and autonomous systems [4]. By combining deep neural networks with reinforcement learning, DRL enables agents to learn complex policies directly from high-dimensional state spaces [17]. However, conventional DRL approaches typically require centralized data collection, which can be impractical or undesirable in distributed and privacy-sensitive environments.

Federated learning (FL) has emerged as a promising solution to train machine-learning models collaboratively without sharing raw data [5], [8]. In FL, multiple agents or clients train local models on their own data and periodically aggregate updates into a global model, preserving data privacy while leveraging collective knowledge. Integrating FL with DRL enables multiple agents to learn collaboratively in a distributed setting, but it introduces challenges, such as non-IID data, communication constraints and unstable policy aggregation [6]-[7].

Recent studies have attempted to address these challenges by applying standard federated averaging (FedAvg) to DRL agents [9]-[10], but performance often degrades in heterogeneous environments due to divergent local updates. Additionally, excessive communication overhead can limit scalability in

---

T.A. Haddad is with LEREESI Laboratory, Higher National School of Renewable Energies, Environment and Sustainable Development, Batna 05078, Algeria. Email: tarek.haddad@hns-re2sd.dz

large-scale multi-agent systems [11]-[12]. These limitations motivate the development of a scalable, stable and communication-efficient federated DRL framework.

Recent advancements in federated reinforcement learning (FedRL) have focused on improving communication efficiency, scalability and learning stability in distributed environments. Di et al. [13] proposed a FedRL-based recommender system that leverages a reinforcement selector and hypernet generator to reduce communication overhead. Zhang et al. [9] introduced a multi-agent approach to optimize federated learning in industrial IoT systems, highlighting the challenges of heterogeneous clients. Pan et al. [11] developed RFCSC, which combines dynamic client selection with adaptive gradient compression for communication-efficient reinforcement learning. Pinto Neto et al. [12] provided a comprehensive survey on FedRL applications in IoT, discussing opportunities and open challenges in privacy-preserving distributed learning. These studies motivate the development of scalable and stable frameworks, like SFDRL, that address both communication and heterogeneity challenges in multi-agent reinforcement learning.

In this paper, we propose Scalable Federated Deep Reinforcement Learning (SFDRL), a collaborative learning framework in which distributed agents perform local DRL training and periodically synchronize with a global model through federated coordination. SFDRL incorporates an adaptive aggregation strategy that weights local updates according to learning performance and stability, as well as a selective participation mechanism that allows only informative agents to communicate, thereby improving scalability and reducing communication overhead in heterogeneous environments. The key contributions of this work are:

- We design an adaptive aggregation mechanism that weights local model updates based on performance and stability, mitigating the effects of non-IID data and unstable learning.
- We introduce selective participation, allowing only agents with significant local improvements to communicate updates, reducing communication overhead while maintaining learning efficiency.
- We provide a comprehensive experimental evaluation in heterogeneous multi-agent environments, demonstrating that SFDRL achieves near-centralized DRL performance with significantly lower communication cost.
- We conduct ablation studies to validate the effectiveness of adaptive aggregation and selective participation in improving stability and scalability.

The remainder of this paper is organized as follows. Section 2 reviews related work on federated reinforcement learning. Section 3 formulates the problem and Section 4 presents the proposed SFDRL algorithm. Sections 5 and 6 provide theoretical analysis and experimental setup, respectively. Section 7 presents results and discussion, including ablation studies. Finally, Section 8 concludes the paper and outlines future research directions.

2. RELATED WORK

Deep Reinforcement Learning (DRL) has achieved significant success in domains, such as robotics, autonomous systems and intelligent transportation [2][3][4]. By combining deep neural networks with reinforcement learning, DRL agents can learn complex policies directly from high-dimensional state spaces. However, conventional DRL typically relies on centralized training and full access to all experience data, which limits its applicability in distributed or privacy-sensitive environments [18].

Federated Learning (FL) enables collaborative training across multiple clients without sharing raw data [5, 8]. In FL, clients train local models and periodically aggregate updates to a global model, preserving privacy while leveraging distributed knowledge. Standard FL approaches, such as FedAvg, face challenges with non-IID data, heterogeneous clients and limited communication bandwidth [7][9]. Truex et al. [16] proposed a hybrid privacy-preserving federated learning approach that combines differential privacy with secure multi-party computation to protect against inference attacks on both exchanged messages and the final model, achieving scalable and accurate training.

Federated Reinforcement Learning (FedRL) integrates DRL and FL to allow multiple agents to learn collaboratively in distributed environments. Early FedRL approaches applied FedAvg to DRL agents, but performance often degrades in heterogeneous settings due to divergent local updates [9][13].

Communication overhead is also a major limitation in large-scale multi-agent systems. Recent advances in federated learning have also explored its application to recommender systems and personalized learning tasks. For example, federated recommender systems have been proposed that leverage diffusion augmentation and guided denoising to enhance recommendation quality under privacy constraints [21].

Recent studies have proposed various strategies to address these challenges. Di et al. [13] introduced a FedRL-based recommender system using a reinforcement selector and hypernet generator to reduce communication. Tian et al. [22] proposed FDDL, a framework that leverages deep reinforcement learning for cache admission and federated learning for parameter sharing, resulting in higher cache hit ratios and lower communication costs compared to conventional and other DRL-based caching schemes. CU-BIC-Learn introduces a reinforcement learning-based enhancement to the CUBIC congestion-control algorithm by using Q-learning to adapt congestion window thresholds based on network feedback [23]. Simulation results show significant improvements in packet loss, bandwidth utilization, latency and fairness compared to standard CUBIC and other classical congestion control schemes.

Communication overhead remains a critical bottleneck in FedRL, particularly for large-scale multi-agent systems. Strategies, such as selective participation, adaptive aggregation and gradient compression have been explored to reduce communication while maintaining learning performance [11], [14]. Zhang et al. [9] demonstrated that multi-agent approaches with optimized client selection can improve both convergence and efficiency in industrial IoT applications. These insights motivate the design of SFDRL, which combines adaptive aggregation and selective participation to achieve near-centralized performance while ensuring scalability and privacy. In addition to communication-efficient strategies, incentive mechanisms for resource-limited devices in federated learning have also been explored. Zhao et al. [15] proposed a learning-based multi-task federated edge learning (FEL) mechanism that jointly designs economic incentives and participation contribution strategies.

In summary, prior work has explored federated learning, reinforcement learning and their integration in distributed and heterogeneous environments. However, existing approaches often either suffer from high communication overhead or reduced learning stability. SFDRL is designed to bridge this gap, providing a scalable, stable and communication-efficient framework for federated multi-agent reinforcement learning.

3. PROBLEM FORMULATION

We consider a collaborative-learning system composed of a set of distributed agents $\mathcal{N} = \{1,2,\dots,N\}$, where each agent interacts with its own local environment and aims to learn an optimal decision-making policy through reinforcement learning. Due to privacy, communication or ownership constraints, raw interaction data cannot be shared among agents. Instead, learning is performed in a federated manner by exchanging model parameters with a coordinating server.

Each agent $i \in \mathcal{N}$ is modeled as a Markov Decision Process (MDP) defined by the tuple $\langle \mathcal{S}_i, \mathcal{A}_i, \mathcal{P}_i, r_i, \gamma \rangle$, where $\mathcal{S}_i$ and $\mathcal{A}_i$ denote the state and action spaces, respectively, $\mathcal{P}_i(s' \mid s, a)$ represents the state-transition probability, $r_i(s,a)$ is the local reward function and $\gamma \in (0,1)$ is the discount factor. The environments may differ across agents, leading to heterogeneous and non-identically distributed (non-IID) data.

Each agent seeks to learn a parameterized policy $\pi_{\theta_i}(a \mid s)$ (or an action-value function $Q_{\theta_i}(s,a)$) that maximizes its expected cumulative discounted return:

$$J_i(\theta_i) = \mathbb{E}_{\pi_{\theta_i}}\left[\sum_{t=0}^{\infty} \gamma^t r_i(s_t, a_t)\right]. \tag{1}$$

In a centralized reinforcement-learning setting, all experiences would be aggregated to optimize a single global model. However, such an approach is impractical in distributed systems due to privacy constraints and communication overhead. To address this, we adopt a federated-learning paradigm in which each agent performs local training and periodically communicates model parameters instead of raw data.

Let θ^t denote the global model parameters at federated-communication round t. At the beginning of each round, the server broadcasts θ^t to a sub-set of participating agents. Each selected agent initializes its local model as $\theta_i^t \leftarrow \theta^t$ and performs E local reinforcement-learning updates through interaction with its environment, producing updated parameters θ_i^{t+1}. The local optimization process can be expressed as:

$$\theta_i^{t+1} = \theta^t - \eta \nabla_\theta \mathcal{L}_i(\theta), \tag{2}$$

where η is the learning rate and $\mathcal{L}_i(\theta)$ denotes the local DRL loss function derived from temporal-difference or policy-gradient updates. Eq. (2) defines the local update of agent i during federated-communication round t. Here, θ_i^{t+1} represents the agent's updated local model parameters after performing E reinforcement-learning steps. The term η denotes the learning rate controlling the step size of each update, while $\nabla_\theta \mathcal{L}_i(\theta)$ is the gradient of the local DRL loss function $\mathcal{L}_i(\theta)$, which can be computed using temporal-difference (TD) or policy-gradient methods depending on the chosen DRL algorithm. This formulation ensures that each agent adjusts its local policy toward minimizing its own expected loss while preserving privacy, as only model parameters - not raw experience data - are communicated to the server. Subsequently, these local updates are aggregated at the central server to refine the global policy θ^{t+1}, which is then broadcast to agents in the next communication round.

After local training, participating agents transmit their updated parameters to the server. The objective of the federated-aggregation process is to compute a global model that reflects the collective learning progress while accounting for heterogeneity and training stability. Formally, the global aggregation can be written as:

$$\theta^{t+1} = \sum_{i \in \mathcal{N}_t} w_i^t \theta_i^{t+1} \tag{3}$$

where $\mathcal{N}_t \subseteq \mathcal{N}$ denotes the set of participating agents at round t and w_i^t represents the aggregation weight associated with agent i, satisfying $\sum_{i \in \mathcal{N}_t} w_i^t = 1$.

Unlike conventional federated averaging, the weights w_i^t are not solely determined by data volume but are designed to reflect the contribution quality of each agent. In particular, they may depend on performance indicators, such as the average episodic return, training stability or improvement magnitude observed during local learning. This formulation allows the global model to emphasize informative and reliable updates while reducing the influence of noisy or unstable ones.

The overall objective of the proposed federated deep reinforcement-learning framework is to learn a global policy parameter vector θ^* that maximizes the aggregated expected return across all agents:

$$\theta^* = \arg\max_\theta \sum_{i \in \mathcal{N}} \mathbb{E}_{\pi_\theta} \left[\sum_{t=0}^{\infty} \gamma^t r_i(s_t, a_t) \right], \tag{4}$$

Subject to decentralized data constraints and limited communication, this formulation captures the fundamental trade-off between collaborative performance, scalability and privacy preservation and serves as the basis for the proposed scalable federated deep reinforcement-learning architecture.

4. PROPOSED METHOD

This section presents the proposed Scalable Federated Deep Reinforcement Learning (SFDRL) method, which enables efficient and robust collaborative learning among distributed agents operating in heterogeneous environments. The proposed approach integrates federated learning with deep reinforcement learning through adaptive-aggregation and selective-participation mechanisms, aiming to improve scalability, stability and learning efficiency under non-IID conditions.

At the beginning of each federated-communication round t, a global model parameterized by θ^t is maintained by a coordinating server. A sub-set of agents $\mathcal{N}_t \subseteq \mathcal{N}$ is selected to participate in the current round. The server broadcasts θ^t to the selected agents, which initialize their local models accordingly. Each agent then interacts with its local environment and performs multiple reinforcement-learning updates using its private experience. The proposed framework is model-agnostic and can be instantiated with either value-based or actor-critic DRL algorithms.

During local training, agent i updates its parameters by minimizing a reinforcement-learning loss function derived from temporal-difference or policy-gradient learning. After E local training episodes, the agent obtains an updated parameter vector θ_i^{t+1} together with performance indicators reflecting the quality of its learning process. These indicators include the average episodic return $\bar{R}_i^t$ and a stability measure σ_i^t, computed as the variance of recent returns. Unlike conventional federated learning, where all clients contribute equally or proportionally to data size, the proposed method evaluates the reliability and usefulness of each update before aggregation.

To address heterogeneity and unstable learning dynamics, an adaptive weighting mechanism is introduced. Each participating agent is assigned a contribution weight w_i^t defined as

$$w_i^t = \frac{\phi\left(\bar{R}_i^t, \sigma_i^t\right)}{\sum_{j \in \mathcal{N}_t} \phi\left(\bar{R}_j^t, \sigma_j^t\right)}, \tag{5}$$

where $\phi(\cdot)$ is a monotonically increasing function with respect to performance and a decreasing function with respect to instability. This design favors agents that exhibit consistent learning progress while reducing the influence of noisy or poorly converged updates. As a result, the aggregation process becomes more robust to non-IIID data distributions and heterogeneous environments.

The global model is updated through a weighted aggregation of local parameters:

$$\theta^{t+1} = \sum_{i \in \mathcal{N}_t} w_i^t \theta_i^{t+1}. \tag{6}$$

This adaptive aggregation enables the global policy to capture shared knowledge across agents while mitigating divergence caused by conflicting local objectives.

To further enhance scalability, the proposed framework incorporates a selective participation mechanism that limits unnecessary communication. Each agent evaluates the significance of its update by measuring the relative improvement in performance between consecutive rounds. Only agents the improvement of which exceeds a predefined threshold ϵ are allowed to transmit their model updates. Formally, agent i participates in round t if:

$$\left|\bar{R}_i^t - \bar{R}_i^{t-1}\right| \geq \epsilon. \tag{7}$$

This mechanism reduces communication overhead and alleviates network congestion, while preserving learning effectiveness by prioritizing informative updates.

The overall training procedure alternates between local reinforcement learning and federated coordination until convergence or a maximum number of communication rounds is reached. Through adaptive aggregation and selective participation, the proposed SFDRL framework achieves improved stability, faster convergence and enhanced scalability compared with standard federated reinforcement-learning approaches. The method preserves data locality and supports heterogeneous agents, making it suitable for large-scale collaborative-learning systems.

4.1 Overall Architecture

The overall architecture of the proposed Scalable Federated Deep Reinforcement Learning (SFDRL) framework is illustrated in Figure 1. The architecture is designed to enable efficient, privacy-preserving and scalable collaborative learning across distributed agents by integrating federated learning, multi-agent reinforcement learning and incentive-aware coordination mechanisms.

The system is composed of three main layers: edge devices, edge servers and a central aggregator. At the edge level, heterogeneous devices (e.g., IoT nodes, smartphones or sensors) perform local interactions with their environments and collect state information. Local training is conducted without sharing raw data, ensuring data privacy. Each device computes local policy updates or state representations and transmits only the necessary model-related information to the upper layer.

At the edge-server level, a multi-agent learning module coordinates the received local information. This layer is responsible for managing multiple agents, handling heterogeneous data distributions and executing collaborative learning through shared representations. An incentive mechanism is integrated to encourage active participation of resource-constrained devices by assigning adaptive rewards based on their contributions. The reward feedback plays a key role in stabilizing training and improving long-term participation. The edge servers also estimate training ratios and intermediate policies that guide local learning behavior.

The central aggregator performs global model aggregation and policy optimization. It collects model updates or policy parameters from edge servers and aggregates them using a federated strategy to produce a global policy. This global policy is then redistributed to the edge servers, closing the learning loop. The aggregation process ensures scalability while reducing communication overhead and preserving privacy.

At the core of the framework lies the proposed SFDRL algorithm, which coordinates reward feedback, aggregation and policy updates across all layers. By jointly optimizing local learning, incentive allocation and global aggregation, SFDRL enables efficient multi-agent collaboration under heterogeneous and communication-constrained environments.

The flowchart in Figure 2 illustrates the main steps of the proposed Scalable Federated Deep Reinforcement Learning (SFDRL) algorithm. The process begins with the initialization of the global policy and agent networks. The global policy is then distributed to edge servers and agents for local training. At the edge-device level, each agent interacts with its environment by observing states, taking actions, receiving rewards and updating its local policy.

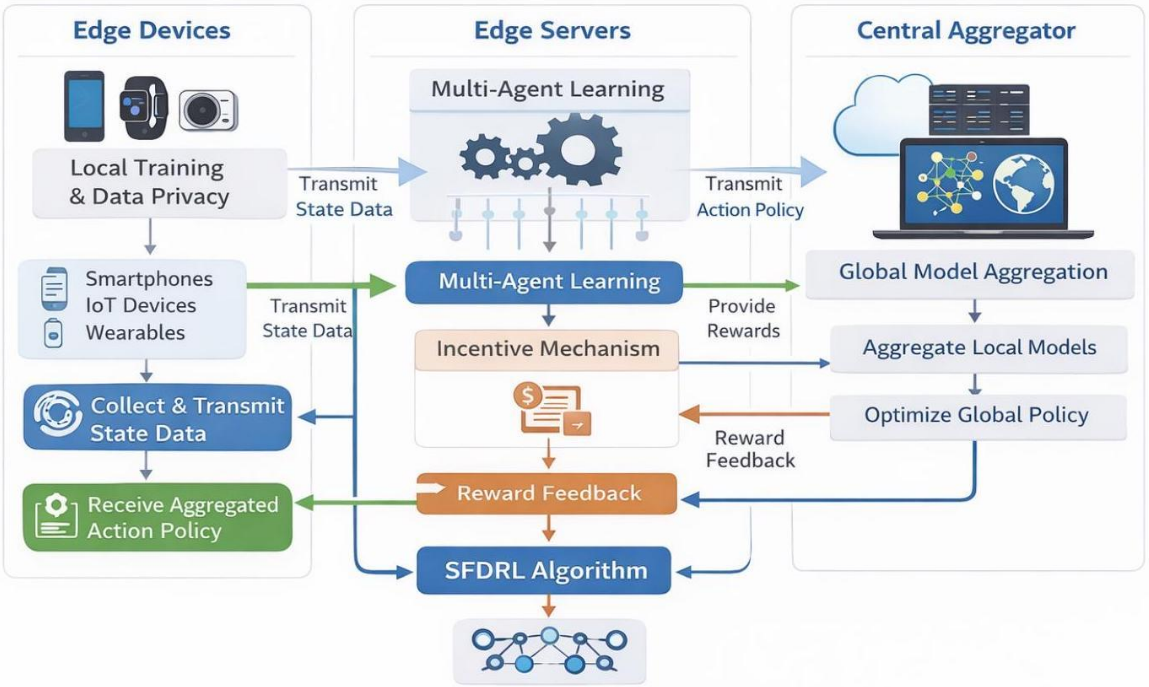


Figure 1. Overall architecture of the proposed SFDRL framework. The system integrates edge devices, edge servers and a central aggregator to enable scalable and privacy-preserving federated deep reinforcement learning with incentive-aware coordination.

Following local training, edge server aggregation collects the local updates from participating devices, optionally applies selective participation and aggregates the results into an edge-level model. The central aggregator then collects edge-level models to update the global policy, which is redistributed to the edge servers, completing the collaborative learning loop. A decision point checks whether the maximum number of episodes is reached or convergence is achieved; if not, the process repeats. This design ensures scalable, communication-efficient and privacy-preserving federated reinforcement learning across heterogeneous multi-agent environments.

4.2 Algorithm Description

The proposed Scalable Federated Deep Reinforcement Learning (SFDRL) algorithm integrates local reinforcement learning with federated coordination to enable collaborative policy optimization across distributed agents while preserving data privacy. At each federated communication round, a sub-set of agents is selected to participate and the current global model parameters are broadcast to them. Each agent initializes its local model with the received parameters and interacts with its environment to collect experience. The local model is updated using standard DRL optimization techniques, such as temporal-difference learning for value-based methods or policy-gradient updates for actor-critic algorithms.

After completing local training episodes, each agent computes performance indicators, including the average episodic return and a stability measure. These indicators are used to determine whether the agent's update is informative enough to contribute to the global model, as governed by the selective participation threshold. Only agents the local updates of which exceed the threshold transmit their parameters to the server, which significantly reduces communication overhead.

The server aggregates the received local updates using an adaptive weighting strategy, in which each agent's contribution is proportional to both the quality and stability of its learning progress. This aggregation produces an updated global model that reflects the collective knowledge of the agents while mitigating the influence of noisy or unstable updates. The global model is then redistributed to the agents in the next round and the process repeats until convergence or until a maximum number of communication rounds is reached.

Overall, the SFDRL algorithm balances exploration and exploitation in local environments, ensures scalability through selective participation and enhances robustness *via* adaptive aggregation. To facilitate adoption, we provide step-by-step instructions for implementing SFDRL. Algorithm 1 summarizes the training loop and provides a step-by-step summary of the complete procedure, while Table 1 provides suggested default hyper-parameters for stable performance across heterogeneous environments.

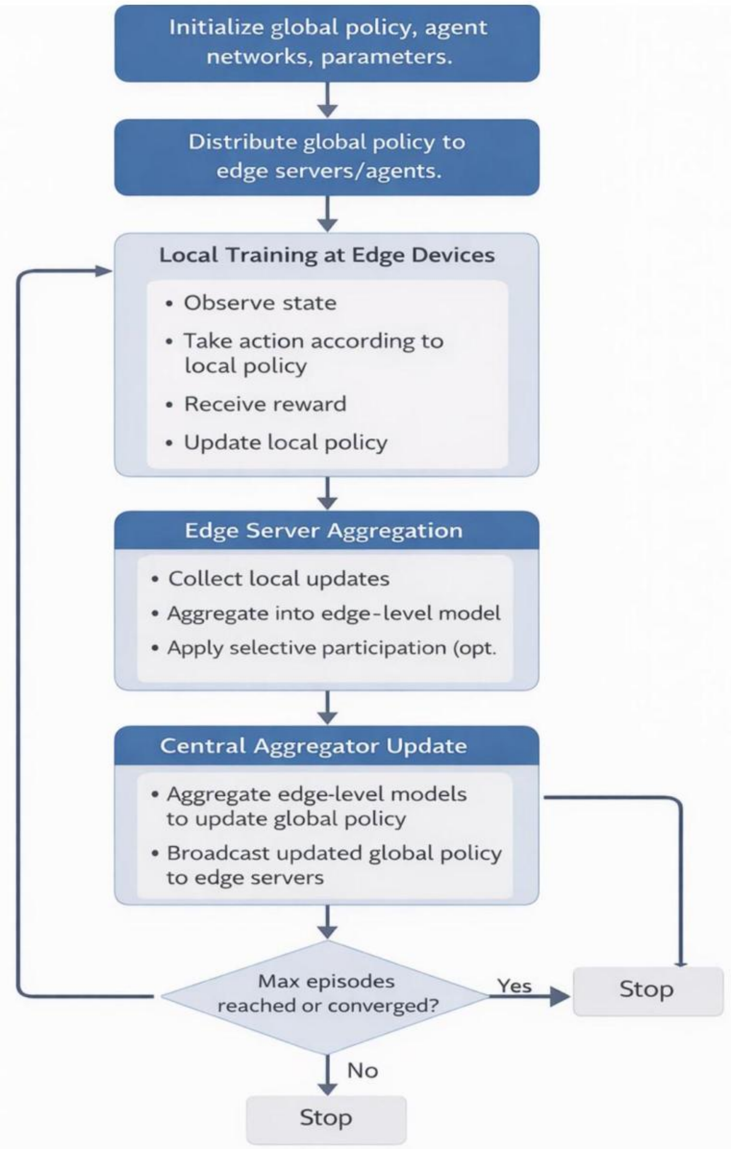
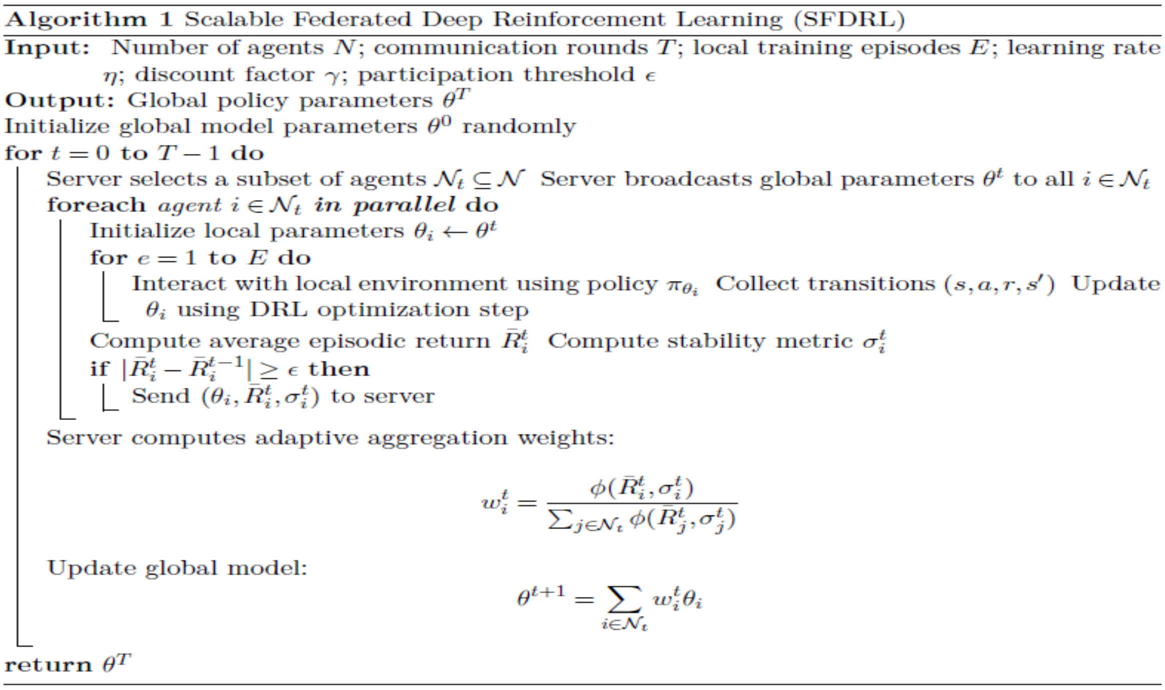

Initialize global policy, agent networks, parameters.
Distribute global policy to edge servers/agents.
Local Training at Edge Devices
• Observe state
• Take action according to local policy
• Receive reward
• Update local policy
Edge Server Aggregation
• Collect local updates
• Aggregate into edge-level model
• Apply selective participation (opt.
Central Aggregator Update
• Aggregate edge-level models to update global policy
• Broadcast updated global policy to edge servers
Max episodes reached or converged?
Yes
Stop
No
Stop

Figure 2. Flowchart of the SFDRL algorithm.

5. THEORETICAL ANALYSIS

In this section, we analyze the theoretical properties of the proposed Scalable Federated Deep Reinforcement Learning (SFDRL) framework, focusing on convergence, stability and communication efficiency in distributed environments. While a formal proof of convergence for general DRL is difficult due to the non-convexity of neural networks and stochastic environment dynamics, we provide intuitive arguments and bounds based on existing federated and reinforcement-learning theory.



```
Algorithm 1 Scalable Federated Deep Reinforcement Learning (SFDRL)
Input: Number of agents $N$; communication rounds $T$; local training episodes $E$; learning rate
       $\eta$; discount factor $\gamma$; participation threshold $\epsilon$
Output: Global policy parameters $\theta^T$
Initialize global model parameters $\theta^0$ randomly
for $t = 0$ to $T - 1$ do
    Server selects a subset of agents $\mathcal{N}_t \subseteq \mathcal{N}$ Server broadcasts global parameters $\theta^t$ to all $i \in \mathcal{N}_t$
    foreach agent $i \in \mathcal{N}_t$ in parallel do
        Initialize local parameters $\theta_i \leftarrow \theta^t$
        for $e = 1$ to $E$ do
            Interact with local environment using policy $\pi_{\theta_i}$ Collect transitions $(s, a, r, s')$ Update
            $\theta_i$ using DRL optimization step
        Compute average episodic return $\bar{R}_i^t$ Compute stability metric $\sigma_i^t$
        if $|\bar{R}_i^t - \bar{R}_i^{t-1}| \geq \epsilon$ then
            Send $(\theta_i, \bar{R}_i^t, \sigma_i^t)$ to server
    Server computes adaptive aggregation weights:
        $$w_i^t = \frac{\phi(\bar{R}_i^t, \sigma_i^t)}{\sum_{j \in \mathcal{N}_t} \phi(\bar{R}_j^t, \sigma_j^t)}$$
    Update global model:
        $$\theta^{t+1} = \sum_{i \in \mathcal{N}_t} w_i^t \theta_i$$
return $\theta^T$
```

5.1 Convergence Intuition

Each agent i performs local reinforcement-learning updates that aim to maximize its expected cumulative return:

$$J_i(\theta_i) = \mathbb{E}_{\pi_{\theta_i}}\left[\sum_{t=0}^{\infty} \gamma^t r_i(s_t, a_t)\right] \tag{8}$$

Under standard DRL assumptions (bounded rewards, learning rate η sufficiently small and sufficient exploration), the local update steps converge to a local optimum of $J_i(\theta_i)$.

The adaptive weighted aggregation at the server ensures that the global model θ^t is a convex combination of stable local updates:

$$\theta^{t+1} = \sum_{i \in \mathcal{N}_t} w_i^t \theta_i^{t+1}, \sum_i w_i^t = 1 \tag{9}$$

By prioritizing updates with high performance and low instability, the aggregation mitigates divergence caused by non-IID environments. Therefore, the global policy progressively approaches a consensus that reflects the collective intelligence of all agents, improving convergence in heterogeneous settings compared with naive federated averaging. Nevertheless, the presented analysis provides a reasonable approximation of the learning dynamics and offers theoretical intuition that is consistent with the empirical results observed in heterogeneous and dynamic experimental settings.

5.2 Stability Analysis

Stability in SFDRL is influenced by two factors: variance in local learning and heterogeneity among agent environments. The weighting function $\phi(\bar{R}_i^t, \sigma_i^t)$ reduces the impact of unstable updates (high σ_i^t) while amplifying reliable contributions. Let $\Delta\theta_i^t$ denote the update magnitude for agent i. The expected deviation of the global model after aggregation can be bounded as:

Table 1. Hyper-parameters for SFDRL implementation.

Parameter	Value	Description
Learning Rate (η)	0.001-0.01	Step size for local DRL updates
Local Updates (E)	10	Number of local training iterations per round
Batch Size	32-128	Number of experiences per gradient update
Discount Factor (γ)	0.99	Weighting of future rewards
Participation Rate (p_t)	0.5-1.0	Fraction of agents participating per round
Aggregation Weighting	Adaptive	Weight local updates based on performance
Communication Rounds (T)	100-500	Total number of federated rounds
Exploration Rate (ϵ)	0.01	ϵ-greedy exploration parameter for DRL

$$\|\theta^{t+1} - \theta^t\| \leq \sum_{i \in \mathcal{N}_t} w_i^t \|\Delta\theta_i^t\| \tag{10}$$

Since unstable updates receive lower weights, large fluctuations are suppressed, leading to a smoother global learning trajectory and improved stability over time.

5.3 Communication Complexity

In standard federated reinforcement learning, all agents communicate updates at every round, leading to high communication costs proportional to $N \cdot T \cdot |\theta|$, where $|\theta|$ is the model size. SFDRL reduces communication *via* selective participation: only agents with meaningful local improvements above a threshold ϵ transmit updates. Denoting p_t as the fraction of participating agents at round t, the total communication complexity becomes:

$$\mathcal{O}\left(|\theta| \sum_{t=1}^{T} p_t N\right), p_t \leq 1, \tag{11}$$

which can be substantially smaller than the naive approach, particularly in large-scale systems with sparse significant updates.

5.4 Discussion

The combination of adaptive aggregation and selective participation provides a theoretical rationale for the observed empirical improvements in convergence and stability. By emphasizing informative updates and suppressing noisy contributions, SFDRL mitigates common challenges in federated reinforcement learning, including non-IID data distributions, heterogeneous agent behaviors and unstable local learning dynamics. Moreover, selective communication ensures scalability without compromising the global learning quality. It is worth noting that the theoretical analysis presented in this section relies on standard assumptions commonly adopted in deep reinforcement learning and federated learning, such as bounded rewards, sufficient exploration and relatively stable local update dynamics. While these assumptions facilitate tractable analysis and provide useful insights into convergence behavior, they may not fully capture the complexity of highly dynamic or non-stationary environments. Extending the theoretical framework to relax these assumptions, for example by explicitly modeling environment dynamics, asynchronous updates or time-varying participation, constitutes an important direction for future work and could further strengthen the robustness of the proposed SFDRL framework.

Overall, the theoretical analysis demonstrates that SFDRL is well-suited for large-scale, distributed and privacy-preserving reinforcement-learning systems.

6. EXPERIMENTAL SETUP

To evaluate the effectiveness of the proposed Scalable Federated Deep Reinforcement Learning (SFDRL) framework, we design experiments that assess convergence, stability, scalability and communication efficiency in distributed environments. The evaluation includes comparisons with centralized DRL and standard federated reinforcement-learning baselines. The chosen simulation-based evaluation allows systematic analysis of scalability and communication efficiency, which would be difficult to isolate in uncontrolled real-world settings.

6.1 Environment and Agents

Experiments are conducted in a set of simulated environments that represent heterogeneous, non-IID scenarios. Each agent $i \in \mathcal{N}$ interacts with a local environment characterized by state space $\mathcal{S}_i$ and action space $\mathcal{A}_i$, as defined in Section 3. Agents receive local rewards $r_i(s, a)$ and update their policies using standard DRL algorithms, including DDPG for continuous action spaces and DQN for discrete action spaces. To assess generality, multiple environments are configured with varying dynamics, stochastic transitions and reward functions.

6.2 Baseline Methods

We compare SFDRL against the following approaches:

- **Centralized DRL** [19]: All agent experiences are aggregated centrally and a single global model is trained. This serves as an upper-bound performance reference, but assumes full data sharing.
- **Independent DRL (IDRL)** [20]: Each agent trains its local DRL model without any collaboration. This highlights the benefits of federated coordination.
- **Federated DRL with Standard FedAvg** [7]: Local DRL models are trained independently and aggregated using conventional federated averaging without adaptive weighting or selective participation.
- **Federated Reinforcement Learning (FedRL)** [13]: Federated Reinforcement Learning (FedRL) is defined as a collaborative learning framework in which multiple agents independently interact with their environments and train local reinforcement-learning models, while a central server periodically aggregates model updates to form a global policy without requiring the sharing of raw data.

In addition to the selected baselines, several recent state-of-the-art federated learning and deep reinforcement-learning methods could be considered for comparison, including communication-efficient federated-optimization techniques, trust-aware or uncertainty-aware aggregation strategies and asynchronous federated reinforcement learning frameworks. However, many of these approaches are designed for supervised or static learning settings or require problem-specific assumptions that make direct and fair comparison with SFDRL challenging. The baselines adopted in this study represent widely used and representative methods in federated and reinforcement learning, providing a meaningful and fair evaluation of the proposed framework.

6.3 Evaluation Metrics

To comprehensively assess the performance of SFDRL and baseline methods, we employ the following metrics:

1) **Cumulative Reward** (R_{cum}): The average episodic return achieved by the global policy across all agents and episodes. Formally,

$$R_{cum} = \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} r_i^t, \tag{12}$$

where N is the number of agents, T is the number of time steps per episode and r_i^t is the reward received by agent i at time t.

2) **Convergence Speed** (C_s): The number of communication rounds required for the global policy to reach a predefined reward threshold R_{th}:

$$C_s = \min\{t: R_{cum}^t \geq R_{th}\}, \tag{13}$$

3) **Stability** (σ^2): Variance of episodic returns over communication rounds:

$$\sigma^2 = \frac{1}{T} \sum_{t=1}^{T} (R_{cum}^t - \bar{R}_{cum})^2, \tag{14}$$

where $\bar{R}_{\text{cum}}$ is the mean cumulative reward over T rounds.

4) **Communication Cost** (C_{comm}) : Total number of model updates transmitted from agents to the server:

$$C_{\text{comm}} = \sum_{t=1}^{T} \sum_{i \in \mathcal{S}_t} \text{size}\left(\theta_i^t\right) \tag{15}$$

where $\mathcal{S}_t$ is the set of agents participating in round t and $\text{size}\left(\theta_i^t\right)$ is the size of the transmitted model.

6.4 Implementation Details

The experiments are implemented in Python using PyTorch. Agents train in parallel on multiple CPU cores and communicate with a central server simulated in the same process. Neural networks for value and policy functions consist of two hidden layers with 128 neurons each, ReLU activation and Adam optimizer with a learning rate $\eta = 0.001$. Discount factor is set to $\gamma = 0.99$ and each communication round consists of $E = 10$ local training episodes. The selective participation threshold ϵ is empirically set to 0.01 to balance communication efficiency and learning performance. Each experiment is repeated 10 times with different random seeds to account for stochasticity.

6.5 Experimental Procedure

At the beginning of each round, the server broadcasts the global model to selected agents. Agents perform local DRL training, compute performance metrics and decide whether to participate in aggregation based on the selective-participation criterion. The server aggregates updates using the adaptive-weighting scheme described in Section 4. The process continues for T communication rounds or until convergence is achieved. All baseline methods follow the same evaluation procedure for a fair comparison.

7. RESULTS AND DISCUSSION

In this section, we present the experimental results of the proposed Scalable Federated Deep Reinforcement Learning (SFDRL) framework and compare its performance with those of baseline methods. We analyze learning efficiency, convergence behavior, stability, scalability and communication efficiency to demonstrate the advantages of our approach.

7.1 Learning Performance

Figure 3 shows the cumulative reward over communication rounds for SFDRL, centralized DRL, independent DRL (IDRL) and standard federated DRL with FedAvg. SFDRL achieves faster convergence and higher final rewards than IDRL and FedAvg, approaching the performance of centralized DRL without sharing raw data. The adaptive aggregation mechanism allows SFDRL to leverage high-quality local updates, leading to improved learning efficiency across heterogeneous agents.

7.2 Convergence and Stability Analysis

Figure 4 summarizes the variance of episodic returns for all methods. SFDRL exhibits lower variance compared with FedAvg and IDRL, demonstrating enhanced stability during training. The adaptive weighting reduces the influence of unstable or poorly performing agents, resulting in smoother global learning trajectories. It is observed that centralized DRL often achieves higher cumulative rewards compared to federated methods, including SFDRL. This performance advantage arises primarily because centralized training has access to the full set of experiences from all agents, enabling the model to learn from the complete state-action distribution. In contrast, federated approaches operate on local data, which may be heterogeneous and non-IID, leading to incomplete or biased learning. Furthermore, centralized DRL updates the model continuously without communication constraints, avoiding delays and aggregation approximations inherent in federated learning. Aggregating local policies in FedAvg or even in SFDRL can introduce inconsistencies when local updates diverge, slightly reducing global-policy performance. Despite this gap, SFDRL narrows the difference by leveraging adaptive aggregation and selective participation, achieving near-centralized performance while maintaining data privacy, reducing communication overhead and enabling scalable multi-agent deployment.

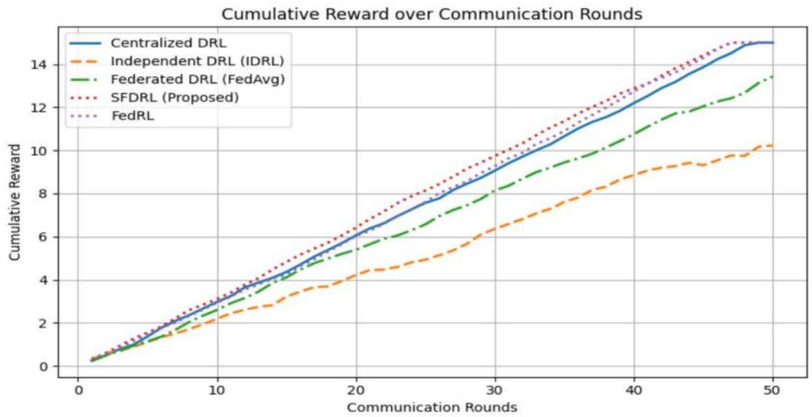


Figure 3. Cumulative reward over federated communication rounds for different methods. SFDRL converges faster and achieves higher returns than baseline federated and independent DRL.

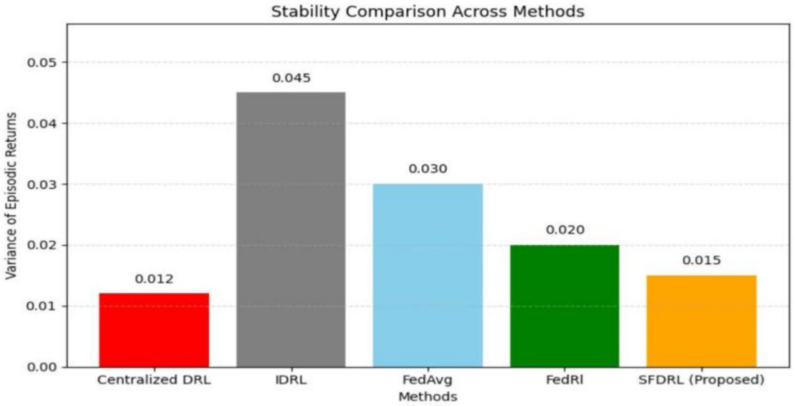


Figure 4. Variance of episodic returns for different methods. Lower variance indicates higher stability.

7.3 Communication Efficiency

Figure 5 presents the total number of transmitted updates during training. SFDRL significantly reduces communication overhead due to selective participation. Only agents with meaningful improvements transmit updates, decreasing redundant transmissions while maintaining performance comparable to full participation federated learning.

7.4 Scalability Analysis

To evaluate scalability, experiments were conducted with varying numbers of agents $N = \{10,20,50,100\}$. SFDRL maintains stable convergence and competitive cumulative rewards as the number of agents increases, whereas FedAvg performance degrades slightly in highly heterogeneous settings. The selective-participation mechanism ensures that only informative updates are aggregated, preventing communication bottlenecks and maintaining learning quality in large-scale deployments.

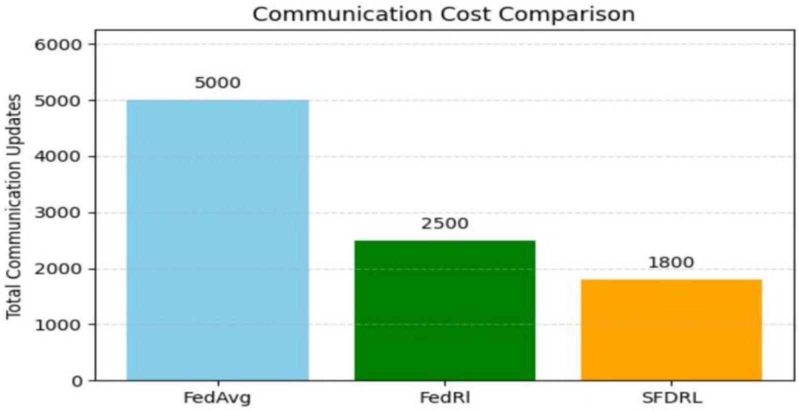


Figure 5. Total communication cost for different federated-learning methods. SFDRL reduces communication overhead while preserving learning efficiency.

7.5 Discussion

The experimental results validate the effectiveness of SFDRL in distributed, heterogeneous environments. Key observations include:

- Adaptive aggregation improves convergence speed and stability by weighting agent contributions according to performance and reliability.
- Selective participation significantly reduces communication costs without sacrificing learning performance.
- SFDRL scales well with the number of agents and is robust to non-IID data distributions.
- Compared with centralized DRL, SFDRL achieves near-optimal performance while preserving data privacy and agent autonomy.

Overall, the results demonstrate that SFDRL provides a practical and efficient framework for large-scale collaborative reinforcement-learning systems. While SFDRL is compared against standard and widely adopted federated and reinforcement-learning baselines, future work will include extensive comparisons with emerging state-of-the-art methods, such as asynchronous and communication-efficient federated DRL frameworks. This will further validate the generality and robustness of SFDRL across diverse learning paradigms.

Although SFDRL preserves data locality by design, potential privacy and security risks remain, as in most federated-learning frameworks. Model updates exchanged during training may leak sensitive information through inference or poisoning attacks. To mitigate these risks, SFDRL can be naturally combined with established privacy-preserving and security mechanisms, such as secure aggregation, differential privacy and robust aggregation strategies. Secure aggregation prevents the server from accessing individual model updates, while differential privacy can be applied to local updates to limit information leakage. In addition, anomaly detection or trust-aware weighting can be incorporated into the adaptive-aggregation process to reduce the impact of malicious or unreliable agents. These extensions are complementary to the proposed framework and represent promising directions for enhancing the privacy and security guarantees of SFDRL.

While the experimental evaluation of SFDRL is conducted in controlled simulated environments, these settings are widely adopted for studying federated reinforcement learning due to their reproducibility and flexibility. Nevertheless, real-world deployment introduces additional challenges, such as unreliable communication, heterogeneous hardware capabilities, delayed updates and non-stationary dynamics.

7.6 Ablation Study

To evaluate the contributions of the key components of SFDRL, we conducted an ablation study by removing one component at a time:

- SFDRL w/o Adaptive Aggregation: All participating agent updates are equally weighted, ignoring performance and stability.
- SFDRL w/o Selective Participation: All selected agents transmit updates regardless of improvement, increasing communication overhead.

Figure 6 summarizes the cumulative reward, convergence speed (number of communication rounds to reach 90% of maximum reward) and total communication cost for each variant.

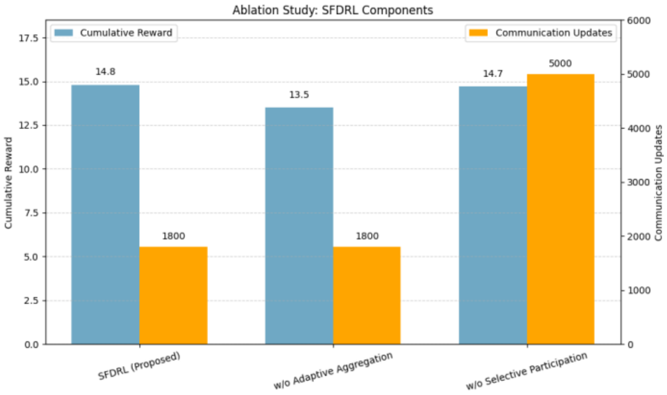


Figure 6. Ablation study results for SFDRL components.

7.7 Discussion

- Removing the adaptive aggregation reduces cumulative reward and slows convergence, indicating that weighting agent contributions according to performance and stability is crucial for effective global learning.
- Removing selective participation increases communication overhead drastically (from 1800 to 5000 updates) with negligible improvement in reward, confirming its role in reducing network load while preserving learning efficiency.
- Together, these results demonstrate that both components are essential for achieving a scalable, stable and communication-efficient federated reinforcement-learning system.

8. CONCLUSION AND FUTURE WORK

In this paper, we proposed a Scalable Federated Deep Reinforcement Learning (SFDRL) framework for collaborative multi-agent learning in heterogeneous and distributed environments. SFDRL integrates adaptive aggregation and selective participation to improve stability, convergence and communication efficiency, achieving near-centralized DRL performance while preserving data privacy. Ablation studies confirmed the importance of both adaptive aggregation and selective participation for effective learning. For future work, we plan to explore more sophisticated aggregation strategies, such as uncertainty-based or trust-aware weighting, extend SFDRL to real-world large-scale applications, like intelligent traffic management and IoT systems, incorporate multi-objective optimization to balance performance, energy efficiency and fairness and investigate advanced privacy-preserving techniques such as differential privacy and secure multiparty computation. Overall, SFDRL provides a scalable, stable and communication-efficient framework that bridges the gap between centralized performance and decentralized, privacy-preserving deployment in collaborative multi-agent systems.

REFERENCES

[1] J. Schulman, F. Wolski, P. Dhariwal, A. Radford and O. Klimov, "Proximal Policy Optimization Algorithms," arXiv preprint, arXiv: 1707.06347, 2017.

[2] V. Mnih et al., "Human-level Control through Deep Reinforcement Learning," Nature, vol. 518, no. 7540, pp. 529-533, Feb. 2015.

[3] L. Li, Y. Lv and F. Wang, "Traffic Signal Timing *via* Deep Reinforcement Learning," IEEE/CAA Journal of Automatica Sinica, vol. 3, no. 3, pp. 247-254, Jul. 2016.

[4] D. Silver et al., "Mastering the Game of Go with Deep Neural Networks and Tree Search," Nature, vol. 529, pp. 484-489, Jan. 2016.

[5] B. McMahan et al., "Communication-efficient Learning of Deep Networks from Decentralized Data," Proc. of the 20th Int. Conf. on Artificial Intelligence and Statistics (AISTATS), vol. 54, pp. 1273-1282, Fort Lauderdale, Florida, USA, 2017.

[6] J. Qi, Q. Zhou, L. Lei and K. Zheng, "Federated Reinforcement Learning: Techniques, Applications and Open Challenges," Intelligence & Robotics, OAE Publishing Inc., DOI: 10.20517/ir.2021.02, 2021.

[7] T. Li, A. K. Sahu, A. Talwalkar and V. Smith, "Federated Learning: Challenges, Methods and Future Directions," IEEE Signal Process. Mag., vol. 37, no. 3, pp. 50-60, May 2020.

[8] Q. Yang, Y. Liu, T. Chen and Y. Tong, "Federated Machine Learning: Concept and Applications," ACM Trans. Intell. Syst. Technol., vol. 10, no. 2, pp. 1-19, Feb. 2019.

[9] W. Zhang et al., "Optimizing Federated Learning in Distributed Industrial IoT: A Multi-agent Approach," IEEE J. Sel. Areas Commun., vol. 39, no. 12, pp. 3688-3703, 2021.

[10] H. Wang, M. Yurochkin, Y. Sun, D. Papailiopoulos and Y. Khazaeni, "Federated Learning with Matched Averaging," arXiv preprint, arXiv: 2002.06440, 2020.

[11] Z. Pan et al., "RFCSC: Communication Efficient Reinforcement Federated Learning with Dynamic Client Selection and Adaptive Gradient Compression," Neurocomputing, vol. 612, p. 128672, 2025.

[12] E. C. Pinto Neto et al., "Federated Reinforcement Learning in IoT: Applications, Opportunities and Open Challenges," Applied Sciences, vol. 13, no. 11, p. 6497, 2023.

[13] Y. Di et al., "FedRL: A Reinforcement Learning Federated Recommender System for Efficient Communication Using Reinforcement Selector and Hypernet Generator," ACM Trans. Recomm. Syst., vol. 4, no. 1, pp. 1-31, 2025.

[14] X. Li, L. Lu, W. Ni, A. Jamalipour, D. Zhang and H. Du, "Federated Multi-agent Deep Reinforcement Learning for Resource Allocation of Vehicle-to-vehicle Communications," IEEE Trans. Veh. Technol., vol. 71, no. 8, pp. 8810-8824, 2022.

[15] N. Zhao et al., "Multi-agent Deep Reinforcement Learning Based Incentive Mechanism for Multi-task Federated Edge Learning," IEEE Trans. Veh. Technol., vol. 72, no. 10, pp. 13530-13535, 2023.

[16] S. Truex et al., "A Hybrid Approach to Privacy-preserving Federated Learning," Proc. of the 12th ACM Workshop on Artificial Intelligence and Security (AISec), pp. 1-11, Nov. 2019.
[17] T. A. Haddad, D. Hedjazi and S. Aouag, "A Deep Reinforcement Learning-based Cooperative Approach for Multi-intersection Traffic Signal Control," Engineering Applications of Artificial Intelligence, vol. 114, p. 105019, 2022.
[18] T. A. Haddad, "Traffic Signal Control for Large-scale Scenario: A Deep Reinforcement Learning-based Cooperative Approach," Proc. of the 12th Int. Conf. Systems and Control (ICSC), pp. 412-417, Batna, Algeria, Nov. 2024.
[19] H. van Hasselt, A. Guez and D. Silver, "Deep Reinforcement Learning with Double Q-learning," Proc. of the 30th AAAI Conf. on Artificial Intelligence (AAAI'16), pp. 2094-2100, 2016.
[20] K. M. Lee et al., "Investigation of Independent Reinforcement Learning Algorithms in Multi-agent Environments," Frontiers in Artificial Intelligence, vol. 5, p. 805823, 2022.
[21] Y. Di et al., "Federated Recommender System Based on Diffusion Augmentation and Guided Denoising," ACM Trans. on Information Systems, vol. 43, no. 2, pp. 1-36, 2025.
[22] A. Tian et al., "Efficient Federated DRL-based Cooperative Caching for Mobile Edge Networks," IEEE Trans. on Network and Service Management, vol. 20, no. 1, pp. 246-260, 2022.
[23] E. Abedini and M. Nickray, "CUBIC-LEARN: A Reinforcement Learning Approach to CUBIC Congestion Control," Jordanian Journal of Computers and Information Technology (JJCIT), vol. 11, no. 4, pp. 466-483, DOI: 10.5455/jjcit.71-1748057293, Dec. 2025.

ملخص البحث:

يمكّن التّعلُّم الاتّحادي من تدريب نماذج التّعلُّم التّعاوني دون مشاركة البيانات الأولية، بينما يوفر التّعلُّم التّعزيزي العميق آليات قوية لاتّخاذ القرارات المتسلسلة. ومع ذلك، يعاني تكاملها من محدودية القابلية للتّوسُّع والحساسية للبيانات غير متطابقة التّوزيع إلى جانب عدم استقرار التّقارب في البيئات الموزّعة. وتقترح هذه الورقة بِنية تعلُّم تعزيزي عميق اتّحادية قابلة للتّوسُّع، حيث تتعلّم العوامل الموزّعة سياساتٍ محلية وتساهم دورياً في نموذج عالمي من خلال استراتيجية تجميع تكيفية تراعي الأداء على العكس من الطّرق التّقليدية التي تعتمد على المتّوسط الموحّد. وتقوم البِنية المقترحة على ترجيح التّحديثات المحلية وفقاً لفعالية تعلّمها، الأمر الّذي يؤدي إلى تقاربٍ أسرع واستقرار أفضل في ظلّ توزيعات البيانات غير المتجانسة، إضافة إلى ذلك، تمّ تقديم آلية اتّصال انتقائية؛ من أجل تقليل عبء الاتّصال بنسبة معتبرة مقارنة بالطُّرق الأخرى.

وقد بينت التّجارب المكثّفة أنّ الطّريقة المقترحة تتفوّق على غيرها من الطّرق، محققةً مكآفاتٍ تراكمية أعلى وتبايناً أقلّ في أثناء التّدريب مع تحسُّنٍ في قابلية التّوسُّع في البيئات الموزّعة كبيرة الحجم. وتؤكد هذه النتائج مُناسبة النّظام المقترح في هذه الدّراسة للتّوظيف العملي في البيئات الذّكية الموزّعة.

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 12, No. 01, March 2026.

AMBIENT BACKSCATTER-ASSISTED PASSIVE RELAYING WITH ENERGY HARVESTING: PERFORMANCE ANALYSIS

Lam-Dong Huynh[1], Lam-Thanh Tu[2], Quang-Sang Nguyen[3] and Tan N. Nguyen[2]

(Received: 11-Dec.-2025, Revised: 1-Feb.-2026, Accepted: 22-Feb.-2026)

ABSTRACT

Ultra-energy-efficient communication solutions are required as Internet of Things (IoT) devices proliferate in the shift to 6G networks. In this paper, a novel architecture that uses energy harvesting (EH) protocols to integrate Ambient Backscatter Communication (AmBC) as a passive relay is investigated. An energy-constrained backscatter device uses a power splitting (PS) mechanism to both reflect its information to the destination and harvest energy for circuit activation. The main contribution of this work is the development of new and accurate closed-form expressions for the system outage probability (OP) over Rayleigh fading channels. Extensive Monte Carlo simulations are conducted to rigorously validate the accuracy of the proposed analytical framework. The analysis reveals important trade-offs between transmission reliability and energy-harvesting efficiency, providing valuable insights for resource optimization in future low-power IoT networks. The results demonstrate that the adverse effects of imperfect successive interference cancellation (SIC) and/or imperfect channel state information (CSI) can be effectively mitigated by increasing the transmit power and/or operating at the optimal value of the reflection coefficient. Moreover, the performance gap between perfect and imperfect SIC and CSI is shown to be relatively small. Finally, we analytically prove that the linear EH model serves as an upper bound for the practical nonlinear EH model.

KEYWORDS

Ambient backscatter communication, Energy harvesting, Outage probability, Relay-assisted network, Imperfect successive interference cancellation, Imperfect channel state information.

1. INTRODUCTION

The proliferation of IoT devices during the transition phase from 5 G to 6 G is driving an escalating demand for ultra-energy-efficient, low-cost communication solutions capable of supporting wide-area connectivity for billions of low-power sensor devices. Recent surveys on the 6 G vision emphasize that future networks must achieve an unprecedented level of dense connectivity, high energy efficiency, and support for near-zero-power IoT nodes [1]-[2]. Furthermore, 6G is being defined as an "Intelligent Network of Everything," where ultra-light, ultra-energy-efficient devices play a critical role in maintaining continuous sensing and communication capabilities [3]. These stringent requirements have spurred intense interest in low-power communication technologies, with AmBC and Wireless Energy Harvesting (WEH) emerging as highly promising solutions due to their ability to reuse ambient signals for communication without requiring an active power source.

To address these connectivity challenges, AmBC has emerged as a particularly critical paradigm, enabling passive devices to communicate by harvesting and reflecting existing radio frequency (RF) signals present in the environment. By not requiring an active RF transmitter, AmBC offers significant advantages in terms of power consumption, cost, and sustainability, making it naturally suited for the development of ultra-energy-efficient IoT systems in the near future [4]. Numerous surveys have comprehensively summarized the principles, system models, challenges, and application directions of AmBC in the field of wireless communication [4]. Beyond the fundamental architectures, extensive research has explored the integration of AmBC with key 6G technologies, such as artificial intelligence (AI), non-terrestrial networks (NTNs), or intelligent platforms, to expand the application space and enhance system performance [5]-[6]. In the context of NTNs, recent studies have further analyzed the

---

1. L.-D. Huynh is with the Faculty of Electrical and Electronics Engineering, Ton Duc Thang University, Ho Chi Minh City, Vietnam. Email: huynhlamdong.st@tdtu.edu.vn
2. L.-T. Tu and Tan N. Nguyen (Corresponding Author) are with the Advanced Intelligent Technology Research Group, Faculty of Electrical and Electronics Engineering, Ton Duc Thang University, Ho Chi Minh City, Vietnam. Emails: tulamthanh@tdtu.edu.vn and nguyennhattan@tdtu.edu.vn
3. Q.-S. Nguyen is with the Posts and Telecommunications Institute of Technology, Ho Chi Minh City, Vietnam. Email: sangnq@ptit.edu.vn

secure performance of satellite-terrestrial networks assisted by backscatter devices to ensure reliability in wide-area IoT deployments [7]. More specialized research directions have also been developed to address the specific challenges of AmBC. For instance, solutions for enhancing coverage and link quality by integrating smart surfaces, like Reconfigurable Intelligent Surfaces (RISs)/Intelligent Reflecting Surfaces (IRSs), have been proposed in several recent works [6][8][9][10]. Meanwhile, studies related to security and privacy focus on analyzing secrecy performance, covert operations, or UAV-assisted AmBC scenarios, thereby assessing the reliability of backscatter systems in complex environments [11][12][13]. In addition, many studies have concentrated on signal detection and decoding techniques, especially in the context where backscatter signals are weak and susceptible to interference from active sources or environmental noise. Methods, such as machine learning-based detection and blind detection, have been developed to improve signal reception efficiency [14]-[15]. Some research also targets hybrid or more advanced architectures, such as LTE-uplink-based AmBC [16], hybrid relay-backscatter systems with energy harvesting [17], NOMA-assisted AmBC, advanced symbiotic radio networks aided by STAR-RIS to maximize IoT throughput [18], or advanced short-packet IoT systems that improve Block Error Rate (BLER) in RIS environments [19] and cooperative/mutualistic AmBC models considering the effects of hardware impairments [20]. Moreover, joint time and energy-management strategies using deep reinforcement learning have been explored in backscatter-assisted hybrid underlay cognitive radio networks (CRNs) to enhance resource efficiency [21]. Furthermore, the performance of device-to-device (D2D) Partial NOMA-assisted Backscatter Communication architectures has been analyzed to evaluate the benefits of combining NOMA and D2D in backscatter systems, particularly in high-density IoT scenarios [22]. Additionally, the impact of co-channel interference on energy harvesting within D2D networks has been investigated, specifically focusing on symbol error-rate analysis under power beacon-assisted configurations [23]. Furthermore, recent advancements in D2D-enabled cellular networks have explored joint throughput maximization and Age of Information (AoI) constraints using deep reinforcement learning to optimize energy-harvesting efficiency [24]. Overall, the current research directions have formed a rich landscape for AmBC, spanning from foundational principles [4], integration with AI and 6 G platforms [5], to the vision of performance enhancement using RISs/IRSs [6], [8]-[10], security solutions [11]-[13], and practical detection-decoding designs and deployment models [10], [13]-[17], [19]-[20].

While AmBC provides an efficient reflection mechanism, the autonomous operation of these devices is fundamentally dependent on a sustainable power supply. Consequently, EH techniques have emerged as a key solution, allowing devices to scavenge energy from the ambient environment (e.g., RF electromagnetic fields, light, heat, or vibration) and convert it into useful electrical power to supply electronic circuits or small storage units. Due to its ability to provide a continuous power source at very low power levels, EH has remained a leading research area in recent years and is gaining increasing attention in subsequent network generations, particularly for ultra-low-power IoT systems [25]-[26]. Specifically, at the hardware and antenna structure level, much research focuses on optimizing the design of antennas, metantennas, and meta-materials to enhance energy-conversion efficiency in the ISM band or other target frequency ranges, thereby increasing the harvested power and improving the efficiency of practical EH systems. Advances in meta-materials/meta-surfaces and metantennas demonstrate the potential for boosting energy capture and increasing deployment flexibility in industrial or civilian scenarios [27]-[29].

Building upon these hardware advancements, at the network layer, EH has been widely integrated into various system designs to support node autonomy: from NOMA-based IoT networks with EH to cooperative relay/DF systems with self-energy recycling mechanisms or UAV-assisted relaying scenarios, where EH plays a pivotal role in balancing reliability and information security. This security-reliability trade-off is also explored through the use of multi-antenna diversity combined with energy harvesting to safeguard wireless communications against eavesdropping [30]. Studies also investigate the trade-off between reliability and security when utilizing EH in relay-assisted UAV communications [31]-[34]. Complementary to these studies, the co-design of clustering, transmission, and trajectory for UAV-assisted Wireless Powered Communication Networks (WPCNs) has been proposed to minimize AoI, ensuring fresh data delivery in energy-constrained environments [35]. Another direction is the combination of EH with intelligent environmental platforms, like RISs-both in the context of energy harvesting for autonomous RISs and in network designs supporting federated learning (FL), where the optimization of RF-EH resource allocation impacts the FL performance. Works on autonomous RISs and EH network optimization for advanced tasks (e.g., federated learning) reflect the trend of tightly integrating communication, distributed processing, and energy management [36]-[37]. Finally, EH is also being considered in practical applications and specific scenarios-from intelligent military logistics based on IoT and EH to physical layer-security solutions utilizing friendly jammers to assist energy-harvesting sensor networks. These results indicate that EH is not merely a hardware-design topic, but also a foundational research axis closely linked to challenges in security, resource allocation optimization and multi-context applications within the 5G/6G ecosystem [38]-[39].

1.1 Motivation and Contribution

The synergy between AmBC and EH represents an ideal strategy for developing ultra-energy-efficient IoT networks. The primary objective of merging these two techniques is to leverage ambient energy to power backscatter devices and utilize supporting relay nodes to extend connectivity range, thereby addressing the needs of massive IoT scenarios. In recent years, the integration of AmBC and EH has garnered significant research interest, leading to a diverse range of studies focusing on different network architectures and optimization goals. Comparative performance studies between backscatter communication and energy harvesting in large-scale IoT networks have elucidated the respective advantages and limitations of each solution, laying the groundwork for hybrid models [40]. Analyses based on Stochastic Geometry (SG) have evaluated the performance of backscatter networks powered by ambient RF energy, providing key metrics, such as outage probability and throughput, in random environments [41]. Complementing these theoretical frameworks, a comprehensive survey has systematically reviewed the practical applications and potential deployment of AmBC in battery-free IoT, emphasizing the necessity of integrating EH to sustain continuous operation [42]. As the field has matured, researchers have begun exploring specialized platforms to further enhance system capability. Specific studies have investigated the use of RISs to simultaneously support backscattering and energy harvesting, thereby improving coverage and enhancing throughput and reliability performance [43]. Additionally, minimizing the Age of Information (AoI) in ambient backscatter-assisted EH-CRNs through cooperative spectrum sensing has emerged as a key approach to maintain data freshness in IoT applications [44]. Meanwhile, hybrid models in cognitive radio networks have been proposed to maximize throughput for backscatter-aided EH networks, demonstrating the feasibility of flexible channel-management mechanisms [45]. In a similar vein, within heterogeneous CRN environments, optimal time-allocation policies have been developed for backscatter-aided relay cooperative transmission to balance harvesting and transmission needs [46]. Further investigations have focused on optimizing transmission protocols and hardware configurations. Opportunistic backscatter protocols have also been developed to optimize energy usage and boost data-transmission performance in EH-assisted IoT networks [47]. Research on relay-assisted cooperative transmission has shown that optimal relay selection in EH-backscatter networks significantly improves throughput and reduces latency, while ensuring continuous operation for battery-free devices [48]. Moreover, the throughput maximization of WPCNs has been further enhanced by employing mobile access points to mitigate the doubly near-far problem and improve energy-transfer efficiency [49]. Optimal control policies for RF-powered backscatter networks have been proposed to balance harvested energy and data-transmission capability [50]. Finally, to address more complex operational requirements, outage analyses in EH-assisted relay networks with backscatter have provided crucial closed-form expressions for system performance evaluation [51]. To accommodate the demands for hybrid long- and short-packet data transmission, numerous works have explored HARQ mechanisms, hybrid packet scheduling, and cooperative relaying for AmBC with EH, aiming to enhance reliability and throughput [52]-[54]. Finally, the combination of EH with cooperative NOMA in two-user backscatter networks has demonstrated the potential to improve communication performance, optimize energy utilization, and address complex IoT scenarios [55].

Despite the recent concentration of research on integrating AmBC with EH [40]-[43], [45], [47]-[48], [50]-[55], a critical gap remains in the literature. Most of these works often focus on throughput or hybrid packet-performance scenarios without a comprehensive analysis of the OP in relay-assisted networks under realistic hardware and channel impairments. For example, studies on opportunistic backscatter [47] or relay selection-based cooperative backscatter [48] primarily aim to maximize throughput without providing closed-form expressions for OP. The hybrid long-short packet models [50]-[52] improve reliability for mixed-packet scenarios, but do not optimally exploit the capabilities of energy harvesting combined with backscatter in a relay environment. Even works concerning cognitive radio networks [45] or D2D Partial NOMA-assisted backscatter [22] do not thoroughly consider the practical constraints of interference cancellation and channel estimation in the outage performance evaluation. In summary, previous methodologies have not simultaneously addressed three key factors: (i) using backscatter for energy efficiency, (ii) exploiting ambient-energy harvesting PS protocol, (iii) analyzing OP performance in a relay-assisted scenario, the impact of imperfect SIC, and the consequences of imperfect CSI due to estimation errors, which are crucial for assessing the practical reliability of IoT networks. To fill this void, our work focuses on ambient backscatter-assisted relaying, with the objective of calculating accurate closed-form expressions for the OP. Compared to previous studies, our work simultaneously combines AmBC and relay-assisted transmission while specifically focusing on OP, constituting a novel, important, and urgent contribution for the deployment of ultra-energy-efficient IoT networks in the 5G/6G era. The main contributions of this paper are summarized as follows:

- We propose a hybrid symbiotic radio framework in which an ambient backscatter device acts as an energy-harvesting-assisted passive relay. Unlike conventional SWIPT systems, the backscatter device is a passive (or semi-passive) node without an active RF transmitter and does not decode the source signal. Instead, it modulates its information by adjusting the reflection coefficient of the incident RF signal. The harvested energy is used solely for circuit activation, while a power-splitting scheme controls the trade-off between harvested energy and reflected signal strength rather than information decoding and energy harvesting.
- We provide a rigorous theoretical analysis of the system performance over Rayleigh fading channels. Specifically, we derive novel closed-form expressions for the OP, taking into account both ideal and imperfect Successive Interference Cancellation (imSIC) mechanisms at the receiver. These expressions, validated by extensive Monte-Carlo simulations, serve as a precise mathematical tool to evaluate the reliability and quantify the symbiotic benefits of the proposed system under residual interference and various channel conditions.
- Our analytical framework's accuracy is rigorously validated through comprehensive Monte-Carlo simulations. The numerical results clearly demonstrate how key parameters - including transmit SNR, target data rates, and channel conditions - directly influence system performance. This thorough investigation provides compelling evidence that our approach reliably captures the essential dynamics affecting overall efficiency and effectiveness.

Table 1. Comparison of the uniqueness of our research to related articles.

Context	**EH protocol**	**EH Modeling**	**Relay-assisted**	**AmBC**	**OP**
Paper [7]	TS	Linear	✓	✓	✓
Paper [21]	TS	Linear		✓	✓
Paper [23]	TS	Linear	✓		✓
Paper [31]	PS	Linear	✓		✓
Paper [32]	TS	Linear	✓		✓
Paper [40]	TS	Linear		✓	
Paper [41]	Hybrid TS - PS	Non - linear		✓	✓
Paper [42]				✓	
Paper [45]	Hybrid Action Selection	Linear		✓	
Paper [46]	TS	Linear	✓	✓	
Paper [47]	TS	Linear		✓	
Paper [48]	TS	Linear	✓	✓	
Paper [50]	TS	Linear		✓	
Paper [52]		Non - linear		✓	✓
Paper [55]	TS	Linear	✓	✓	✓
This paper	PS	Linear and Non-linear	✓	✓	✓

The remainder of the paper is organized as follows. Section 2 gives an overview of the system model. Section 3 presents the information-theoretic mathematical framework, to achieve the performance of the system. Section 4 presents numerical results and discussion to validate the developed framework as well as deeply explore the impacts of system key parameters, while Section 5 provides concluding remarks.

2. SYSTEM MODEL

Figure 1 illustrates the proposed hybrid wireless communication system model, which has been widely investigated, set within a dense urban environment. The system comprises three main entities: a high-power Source (S), an energy-constrained Backscatter device (B), and an end Device (D)[1]. S serves as the primary access point, simultaneously transmitting the RF signal for energy provisioning and the data signal. B operates by harvesting RF energy from S *via* the Energy-harvesting link (green line). Concurrently, B modulates its own information onto the incident RF signal and reflects it towards D, forming one component of the Information links (black lines). The ultimate receiver, D, collects information through a combination of the conventional direct link from S and the backscatter link from B. This model is designed to thoroughly investigate the trade-off and co-existence between energy efficiency (enabled by B) and overall communication performance in the system. The mechanism for signal distribution and energy provision is realized through the PS scheme at S, with a total transmit power P_S over a transmission-block duration T. The system utilizes a reflection coefficient $\beta(0 \leq \beta \leq 1)$ to divide P_{S} into two streams supporting the backscatter link: a fraction $(1-\beta)P_{\mathrm{S}}$ is allocated for the energy-harvesting process at B to power the device, while the remaining fraction βP_{S} is used to generate the RF backscattering signal, which acts as the carrier wave for B's data modulation. This modulated signal is then reflected by B towards D. Concurrently, the conventional direct communication link from S to D is maintained with a constant power P_{S}. D collects information through a combination of the direct link and the backscatter link. This PS scheme is crucial, as it enables the system to control the critical trade-off between the harvested energy at B and the received power of the backscattered signal at D, thereby optimizing the overall system performance and energy efficiency.

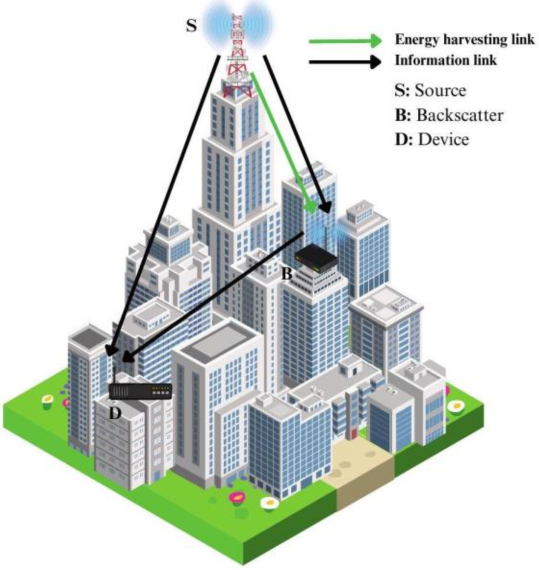


Figure 1. System model.

Let us denote $h_{\mathrm{SD}}, h_{\mathrm{SB}}, h_{\mathrm{BD}}$ as channel coefficients of the direct link from source node S to destination node D, and $\mathrm{S} \to \mathrm{B}, \mathrm{B} \to \mathrm{D}$ links, respectively. We assume Rayleigh fading channels. Channel gains $h_{\mathrm{X}}, \mathrm{X} \in \{\mathrm{SB}, \mathrm{BD}, \mathrm{SD}\}$ are exponential random variables (RVs) the probability density function (PDF) and cumulative distribution function (CDF) of which are given as, respectively.

$$f_{\gamma_{\mathrm{X}}}(x) = \lambda_{\mathrm{X}} \exp(-\lambda_{\mathrm{X}} x)$$
$$F_{\gamma_{\mathrm{X}}}(x) = 1 - \exp(-\lambda_X x) \tag{1}$$

where λ_{X} is mean of RV X. To take into account the simple path-loss model, we can formulate the parameters as follows:

$$\lambda_{\mathrm{X}} = (d_{\mathrm{X}})^{\omega}, \tag{2}$$

---

[1] Although the considered AmBC system involves a single backscatter device equipped with a single antenna, our work explicitly incorporates energy harvesting with an energy-sufficiency constraint and derives closed-form expressions to characterize whether the harvested energy is sufficient for backscatter operation and signal decoding, thereby providing insights not captured in existing studies.

where ω is the path-loss exponent and d_{X} is the distance between two respective nodes. In a practical linear energy-harvesting paradigm, the average received power P_{h} at the BD is expressed as:

$$P_{\mathrm{h}} = (1-\beta)P_{\mathrm{S}}\gamma_{\mathrm{SB}}. \tag{3}$$

If $P_{\mathrm{h}} < P_{\mathrm{th}}$, then B (Backscatter device) does not have enough energy to backscatter the signal from S to D. Thus, the received signal at D is written as: where P_{th} is the saturation threshold of the rechargeable power of the hardware circuit.

$$y_{\mathrm{D}}^{\mathrm{I}} = \sqrt{P_{\mathrm{S}}}h_{\mathrm{SD}}x_{\mathrm{S}} + n_{\mathrm{D}}^{\mathrm{I}}. \tag{4}$$

In this case, the received SNR at D required to successfully decode x_{S} is calculated as:

$$\gamma_{\mathrm{D}}^{\mathrm{I}} = \frac{|h_{\mathrm{SD}}|^2 P_{\mathrm{S}}}{N_0} = \Psi\gamma_{\mathrm{SD}} \tag{5}$$

where $\Psi = \frac{P_{\mathrm{S}}}{N_0}$ is the average transmit power to noise ratio. If $P_{\mathrm{h}} \geq P_{\mathrm{th}}$, then B (Backscatter device) has enough energy to backscatter the signal to D. The received signal at D is then described by:

$$y_{\mathrm{D}}^{\mathrm{II}} = \sqrt{P_{\mathrm{S}}}h_{\mathrm{SD}}x_{\mathrm{S}} + h_{\mathrm{BD}}h_{\mathrm{SB}}\sqrt{\beta P_{\mathrm{S}}}x_{\mathrm{B}}x_{\mathrm{S}} + n_{\mathrm{D}}^{\mathrm{II}} \tag{6}$$

with $\mathbb{E}\{|x_{\mathrm{B}}|^2\} = 1$.

The destination D first decodes the signal x_{S} by treating the backscattered signal as interference, with the received SINR (Signal-to-Interference-plus-Noise Ratio) given by:

$$\gamma_{\mathrm{S}} = \frac{|h_{\mathrm{SD}}|^2 P_{\mathrm{S}}}{\beta P_{\mathrm{S}}|h_{\mathrm{BD}}|^2|h_{\mathrm{SB}}|^2 + N_0} = \frac{|h_{\mathrm{SD}}|^2\Psi}{\beta\Psi|h_{\mathrm{BD}}|^2|h_{\mathrm{SB}}|^2 + 1} \approx \frac{|h_{\mathrm{SD}}|^2}{\beta|h_{\mathrm{BD}}|^2|h_{\mathrm{SB}}|^2} = \frac{\gamma_{\mathrm{SD}}}{\beta\gamma_{\mathrm{SB}}\gamma_{\mathrm{BD}}} \tag{7}$$

Next, leveraging the SIC technique, x_{S} is subtracted from the composite received signal, i.e., $y_{\mathrm{D}}^{\mathrm{III}} = y_{\mathrm{D}}^{\mathrm{II}} - \sqrt{P_{\mathrm{S}}}h_{\mathrm{SD}}x_{\mathrm{S}} + n_{\mathrm{D}}^{\mathrm{III}}$, here $n_{\mathrm{D}}^{\mathrm{III}}$ denotes the AWGN introduced during the SIC operation, D can then decode the backscatter message x_{B} from B . Hence, the SNR of decoding x_{B} at D can be expressed as:

$$\gamma_{\mathrm{B}} = \beta\Psi|h_{\mathrm{SB}}|^2|h_{\mathrm{BD}}|^2 = \beta\Psi\gamma_{\mathrm{SB}}\gamma_{\mathrm{BD}} \tag{8}$$

Finally, the backscatter signal can be successfully decoded when x_{S} and x_{B} are perfectly decoded at D. Thus, based on Equations (7) and (8), the end-to-end received SINR and SNR at D can be claimed by:

$$\gamma_{\mathrm{D}}^{\mathrm{II}} = \min(\gamma_{\mathrm{S}}, \gamma_{\mathrm{B}}) = \min\left(\frac{\gamma_{\mathrm{SD}}}{\beta\gamma_{\mathrm{SB}}\gamma_{\mathrm{BD}}}, \beta\Psi\gamma_{\mathrm{SB}}\gamma_{\mathrm{BD}}\right) \tag{9}$$

3. PERFORMANCE ANALYSIS

3.1 Outage Probability Analysis

In this section, we provide a detailed and formal definition of the OP to evaluate the communication reliability of the proposed symbiotic radio system. An outage event occurs when the destination D is unable to successfully decode the intended information. Given the hardware constraints of the energy-constrained backscatter device B, the system performance is analyzed based on two mutually exclusive states of EH.

The first state represents a scenario where the backscatter device fails to harvest sufficient energy to activate its circuits ($P_{\mathrm{h}} < P_{\mathrm{th}}$). In this case, the system can only utilize the direct link from the source S. The second state occurs when B harvests enough energy ($P_{\mathrm{h}} \geq P_{\mathrm{th}}$), allowing it to function as a passive relay. In this symbiotic mode, D leverages the SIC technique to decode both primary and backscatter signals.

Formally, the OP of the proposed system is defined as the sum of the probabilities of these two independent events:

$$\mathrm{OP} = \Pr\left(P_{\mathrm{h}} < P_{\mathrm{th}}, \gamma_{\mathrm{D}}^{\mathrm{I}} < \gamma_{\mathrm{th}}\right) + \Pr\left(P_{\mathrm{h}} \geq P_{\mathrm{th}}, \gamma_{\mathrm{D}}^{\mathrm{II}} < \gamma_{\mathrm{th}}\right) \tag{10}$$

where $\gamma_{\text{th}} = 2^{\text{R}} - 1$ is the threshold of the system and R is the target rate of the source. By employing the concept of normalized bandwidth ($\text{B} = 1$ Hz), the target transmission rate R is defined as the spectral efficiency in bits/s/Hz. According to the Shannon-Hartley theorem, the minimum SINR required to support this rate, defined as the system threshold γ_{th}, is obtained by solving $R = \log_2(1+\gamma_{\text{th}})$, which yields the relation $\gamma_{\text{th}} = 2^{\text{R}} - 1$. This formulation provides a formal physical insight into the system's operation, where the overall reliability is modeled as a weighted sum of two independent energy states. It reveals that the system performance is fundamentally constrained by a critical coupling between the hardware-level energy-harvesting threshold P_{th} and the information theoretic-decoding requirement γ_{th}, characterizing the operational limits of battery-free backscatter-assisted relaying. By substituting the expressions for $\gamma_{\text{D}}^{\text{I}}$ (from Equation 5) and $\gamma_{\text{D}}^{\text{II}}$ (from equation (9) into equation (10), we gather the final expression for the outage probability:

$$\text{OP} = \underbrace{\Pr\left\{\gamma_{\text{SB}} < \frac{P_{\text{th}}}{\Psi(1-\beta)}, \Psi\gamma_{\text{SD}} < \gamma_{\text{th}}\right\}}_{P_1} + \underbrace{\Pr\left\{\gamma_{\text{SB}} \geq \frac{P_{\text{th}}}{\Psi(1-\beta)}, \min\left(\frac{\gamma_{\text{SD}}}{\beta\gamma_{\text{SB}}\gamma_{\text{BD}}}, \beta\Psi\gamma_{\text{SB}}\gamma_{\text{BD}}\right) < \gamma_{\text{th}}\right\}}_{P_2} \quad (11)$$

where $\Phi = \frac{P_{\text{th}}}{N_0}$.

From (11), P_2 is calculated as:

$$P_2 = \begin{cases} \Pr\left(\min\left(\frac{\gamma_{\text{SD}}}{\beta x\gamma_{\text{BD}}}, \beta\Psi x\gamma_{\text{BD}}\right) < \gamma_{\text{th}}\right) & , x \geqslant \frac{\Phi}{(1-\beta)\Psi} \\ 0 & , x < \frac{\Phi}{(1-\beta)\Psi} \end{cases}$$

$$= \int_{\frac{\Phi}{(1-\beta)\Psi}}^{+\infty} \underbrace{\Pr\left(\min\left(\frac{\gamma_{\text{SD}}}{\beta x\gamma_{\text{BD}}}, \beta\Psi x\gamma_{\text{BD}}\right) < \gamma_{\text{th}}\right)}_{\Theta} f_{\gamma_{\text{SB}}}(x)dx \quad (12)$$

From (12), Θ can be calculated as:

$$\Theta = 1 - \Pr\left(\min\left(\frac{\gamma_{\text{SD}}}{\beta x\gamma_{\text{BD}}}, \beta\Psi x\gamma_{\text{BD}}\right) \geq \gamma_{\text{th}}\right) \overset{(a)}{=} 1 - \Pr\left(\frac{\gamma_{\text{SD}}}{\beta x\gamma_{\text{BD}}} \geq \gamma_{\text{th}}, \beta\Psi x\gamma_{\text{BD}} \geq \gamma_{\text{th}}\right)$$

$$= 1 - \Pr\left(\frac{\gamma_{\text{th}}}{\beta\Psi x} \leq \gamma_{\text{BD}} \leq \frac{\gamma_{\text{SD}}}{\beta x\gamma_{\text{th}}}\right) = 1 - \int_0^{+\infty} \left[F_{\gamma_{\text{BD}}}\left(\frac{y}{\beta x\gamma_{\text{th}}}\right) - F_{\gamma_{\text{BD}}}\left(\frac{\gamma_{\text{th}}}{\beta\Psi x}\right)\right] f_{\gamma_{\text{SD}}}(y)dy \quad (13)$$

where step (a) is obtained by applying the property of the minimum of two random variables; i.e., $\Pr(\min(A,B) \geq \gamma_{\text{th}}) = \Pr(A \geq \gamma_{\text{th}}, B \geq \gamma_{\text{th}})$. By substituting Eq. (1) into Eq. (13) and performing several algebraic manipulations, the closed-form expression for Θ can be obtained as follows:

$$\Theta = 1 - \int_0^{+\infty} \left[\exp\left(-\frac{\lambda_{\text{BD}}\gamma_{\text{th}}}{\beta\Psi x}\right) - \exp\left(-\frac{\lambda_{\text{BD}}y}{\beta x\gamma_{\text{th}}}\right)\right] f_{\gamma_{\text{SD}}}(y)dy$$

$$= 1 - \exp\left(-\frac{\lambda_{\text{BD}}\gamma_{\text{th}}}{\beta\Psi x}\right) + \lambda_{\text{SD}}\int_0^{+\infty} \exp\left(-\frac{\lambda_{\text{BD}}y}{\beta x\gamma_{\text{th}}} - \lambda_{\text{SD}}y\right)dy$$

$$= 1 - \exp\left(-\frac{\lambda_{\text{BD}}\gamma_{\text{th}}}{\beta\Psi x}\right) + \frac{\lambda_{\text{SD}}\beta x\gamma_{\text{th}}}{\lambda_{\text{BD}} + \lambda_{\text{SD}}\beta x\gamma_{\text{th}}} \quad (14)$$

Substituting Eq. (14) into Eq. (12), we have:

$$P_2 = \int_{\frac{\Phi}{(1-\beta)\Psi}}^{+\infty} \left(1 - \exp\left(-\frac{\lambda_{\text{BD}}\gamma_{\text{th}}}{\beta\Psi x}\right) + \frac{\lambda_{\text{SD}}\beta x\gamma_{\text{th}}}{\lambda_{\text{BD}} + \lambda_{\text{SD}}\beta x\gamma_{\text{th}}}\right)\lambda_{\text{SB}}\exp(-\lambda_{\text{SB}}x)dx$$

$$= \exp\left(-\frac{\lambda_{\text{SB}}\Phi}{(1-\beta)\Psi}\right) + \lambda_{\text{SB}}\int_{\frac{\Phi}{(1-\beta)\Psi}}^{+\infty} \left(\frac{\beta x\gamma_{\text{th}}\lambda_{\text{SD}}}{\lambda_{\text{BD}} + \beta x\gamma_{\text{th}}\lambda_{\text{SD}}} - \exp\left(-\frac{\lambda_{\text{BD}}\gamma_{\text{th}}}{\beta\Psi x}\right)\right)\exp(-\lambda_{SB}x)dx. \quad (15)$$

However, the integral in Equation (15) is a tough task to find a closed-form expression. Hence, by applying the Gaussian-Chebyshev quadrature in [38], SOP can be approximated. As a result, with

$\phi_n = \cos\left(\frac{2n-1}{2N}\pi\right)$, P_2 can be reformulated as:

$$
\begin{aligned}
P_2 &\approx \exp\left(-\frac{\lambda_{\mathrm{SB}}\Phi}{(1-\beta)\Psi}\right) + \frac{\lambda_{\mathrm{SB}}\pi^2}{4N}\sum_{n=1}^{N} \sqrt{1-\varphi_n^2}F\left(\tan\left((\varphi_n+1)\frac{\pi}{4}\right) + \frac{\Phi}{(1-\beta)\Psi}\right)\sec^2\left((\varphi_n+1)\frac{\pi}{4}\right) \\
&= \exp\left(-\frac{\lambda_{\mathrm{SB}}\Phi}{(1-\beta)\Psi}\right) + \frac{\lambda_{\mathrm{SB}}\pi^2}{4N}\sum_{n=1}^{N} \left(\begin{array}{c} \sqrt{1-\varphi_n^2}\left(\begin{array}{c} \frac{\beta\left(\tan\left((\varphi_n+1)\frac{\pi}{4}\right)+\frac{\Phi}{(1-\beta)\Psi}\right)\gamma_{\mathrm{th}}\lambda_{\mathrm{SD}}}{\lambda_{\mathrm{BD}}+\beta\left(\tan\left((\varphi_n+1)\frac{\pi}{4}\right)+\frac{\Phi}{(1-\beta)\Psi}\right)\gamma_{\mathrm{th}}\lambda_{\mathrm{SD}}} \\ -\exp\left(-\frac{\lambda_{\mathrm{BD}}\gamma_{\mathrm{th}}}{\beta\Psi\left(\tan\left((\varphi_n+1)\frac{\pi}{4}\right)+\frac{\Phi}{(1-\beta)\Psi}\right)}\right) \end{array}\right) \\ \times \exp\left(-\lambda_{\mathrm{SB}}\left(\tan\left((\varphi_n+1)\frac{\pi}{4}\right)+\frac{\Phi}{(1-\beta)\Psi}\right)\right)\sec^2\left((\varphi_n+1)\frac{\pi}{4}\right) \end{array}\right)
\end{aligned}
\quad (16)
$$

Substituting Eq. (1) into Eq. (11) yields the calculation of P_1 as follows:

$$
P_1 = \Pr\left(\gamma_{\mathrm{SB}} < \frac{\Phi}{(1-\beta)\Psi}, \gamma_{\mathrm{SD}} < \frac{\gamma_{\mathrm{th}}}{\Psi}\right) = \left(1-\exp\left(-\frac{\lambda_{\mathrm{SB}}\Phi}{(1-\beta)\Psi}\right)\right)\left(1-\exp\left(-\frac{\lambda_{\mathrm{SD}}\gamma_{\mathrm{th}}}{\Psi}\right)\right) \quad (17)
$$

Substituting Eq. (16) and Eq. (17) into Eq. (11), we have:

$$
\begin{aligned}
\mathrm{OP} &= \left(1-\exp\left(-\frac{\lambda_{\mathrm{SB}}\Phi}{(1-\beta)\Psi}\right)\right)\left(1-\exp\left(-\frac{\lambda_{\mathrm{SD}}\gamma_{\mathrm{th}}}{\Psi}\right)\right) + \exp\left(-\frac{\lambda_{\mathrm{SB}}\Phi}{(1-\beta)\Psi}\right) \\
&\quad + \frac{\lambda_{\mathrm{SB}}\pi^2}{4N}\sum_{n=1}^{N} \left(\begin{array}{c} \sqrt{1-\varphi_n^2}\left(\begin{array}{c} \frac{\beta\left(\tan\left((\varphi_n+1)\frac{\pi}{4}\right)+\frac{\Phi}{(1-\beta)\Psi}\right)\gamma_{\mathrm{th}}\lambda_{\mathrm{SD}}}{\lambda_{\mathrm{BD}}+\beta\left(\tan\left((\varphi_n+1)\frac{\pi}{4}\right)+\frac{\Phi}{(1-\beta)\Psi}\right)\gamma_{\mathrm{th}}\lambda_{\mathrm{SD}}} \\ -\exp\left(-\frac{\lambda_{\mathrm{BD}}\gamma_{\mathrm{th}}}{\beta\Psi\left(\tan\left((\varphi_n+1)\frac{\pi}{4}\right)+\frac{\Phi}{(1-\beta)\Psi}\right)}\right) \end{array}\right) \\ \times \exp\left(-\lambda_{\mathrm{SB}}\left(\tan\left((\varphi_n+1)\frac{\pi}{4}\right)+\frac{\Phi}{(1-\beta)\Psi}\right)\right)\sec^2\left((\varphi_n+1)\frac{\pi}{4}\right) \end{array}\right)
\end{aligned}
\quad (18)
$$

The closed-form expression in Equation (18) reveals several key mathematical insights into the system's behavior. First, the power-splitting factor β acts as a "mathematical pivot" that governs the trade-off between the activation probability of the backscatter device, which depends on $(1-\beta)$ and the energy threshold Φ, and the resulting signal quality at the destination, which is proportional to β. Second, the analytical structure highlights a "product-of-exponentials" characteristic arising from the cascaded channel gain $|h_{\mathrm{SB}}|^2|h_{\mathrm{BD}}|^2$. This indicates that a deep fade in either the forward link (S → B) or the backscatter link (B → D) will significantly dominate the overall outage performance. Finally, as the transmit SNR (Ψ) increases, the OP does not vanish to zero, but instead converges to a fixed performance floor. This demonstrates that the reliability of battery-free IoT devices is fundamentally constrained by the tight coupling between energy harvesting requirements (Φ) and the actual channel conditions.

4. NUMERICAL RESULTS AND SIMULATIONS

In this section, we provide numerical results to not only verify the accuracy of the proposed mathematical frameworks, but also discuss the behaviors of the considered systems under the impact of various important parameters by using the Monte-Carlo approach. To ensure the validity and practical relevance of our analysis, we used the simulation parameters such as those in [56]-[60]. These parameters are summarized in Table 2.

Figure 2 illustrates the OP *versus* the average transmit Signal-to-Noise Ratio (Ψ) in dB, considering various required data rates ($R = \{0.15, 0.25, 0.5\}\text{bps/Hz}$). The close agreement between the theoretical analysis curves and the Monte-Carlo simulation results validates the accuracy of the closed-form outage probability expressions derived in the mathematical-analysis section. First, a monotonic decreasing trend in OP is observed as Ψ increases from 0 dB to 30 dB across all scenarios. This can be attributed to the fact that an increase in Ψ corresponds to higher transmit power P_S, which enables B to harvest sufficient energy to overcome the circuit-activation threshold (P_{th}). Simultaneously, the powers of both the reflected and direct signals at D are enhanced, thereby improving the received SINR and reducing the system-outage probability. Second, the required data rate R has a significant negative impact on the system performance. Specifically, at a fixed transmit power level (e.g., $\Psi = 10$ dB), as R increases from 0.15bps/Hz to 0.5bps/Hz, the OP rises sharply (indicated by the blue curve lying significantly higher than the red curve). This is because the SNR decoding threshold, defined as $\gamma_{\text{th}} = 2^R - 1$, increases with R. A higher γ_{th} necessitates superior channel conditions and sufficient harvested energy for successful decoding, consequently leading to a higher probability of outage events.

Table 2. Simulation parameters.

Symbol	**Parameter name**	**Value**
Ψ	Transmit power to noise ratio at S	0 to 30(dB)
Φ	Threshold of the transmit power to noise ratio at B	-10 to 5(dB)
R	Target rate	0.15 to 0.5bps/Hz
λ_{SB}	Mean of $\lvert h_{\text{SB}}\rvert^2$	0.25 to 4
λ_{BD}	Mean of $\lvert h_{\text{BD}}\rvert^2$	0.25 to 4
λ_{SD}	Mean of $\lvert h_{\text{SD}}\rvert^2$	0.25 to 2
N	The Gauss Chevbyshev parameter	80
β	The power-splitting factor	0 to 1

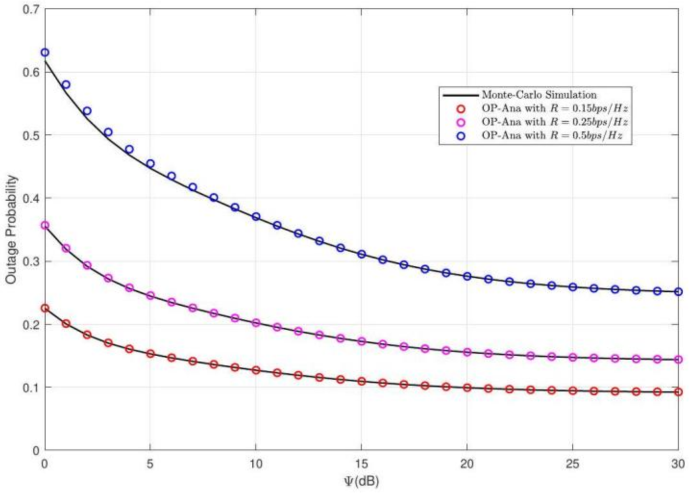


Figure 2. OP *versus* Ψ(dB), with varying R .

Figure 3 provides a detailed insight into how system characteristics affect the OP, with the distinction between linear and nonlinear models being prominently displayed. The adopted nonlinear energy-harvesting model is given below [61]-[62]

$$P_{\text{nEH}} = \frac{P_{\text{B}}^{\max}\left(1 - \exp(-\epsilon_1 P_{\text{h}} + \epsilon_1 \epsilon_0)\right)}{1 + \exp(-\epsilon_1 P_{\text{h}} + \epsilon_1 \epsilon_2)}, \tag{19}$$

where ϵ_0 represents the sensitivity threshold, ϵ_1 is the resistance parameter, and ϵ_2 denotes the capacitance parameter of the harvesting circuit; $P_{\text{B}}^{\max}$ is the maximum output supported by the circuit. Specifically, curves employing linear models consistently achieve superior performance, maintaining a significantly lower OP than their nonlinear counterparts, which suffer from signal distortion and component saturation. Similarly, imperfect SIC circumstances result in a larger OP because of residual interference from partially canceled signals. However, in these real-world situations, raising the average transmit power to noise ratio, Ψ, is a useful tactic to improve the OP and boost overall system performance in the low-power regime. The graph, however, also draws attention to a critical physical limit: this power enhancement is only effective up to a certain threshold; as Ψ increases, the OP eventually enters an error floor brought on by dominant residual interference and nonlinearities, meaning that additional power increments no longer produce the notable OP reductions seen initially.

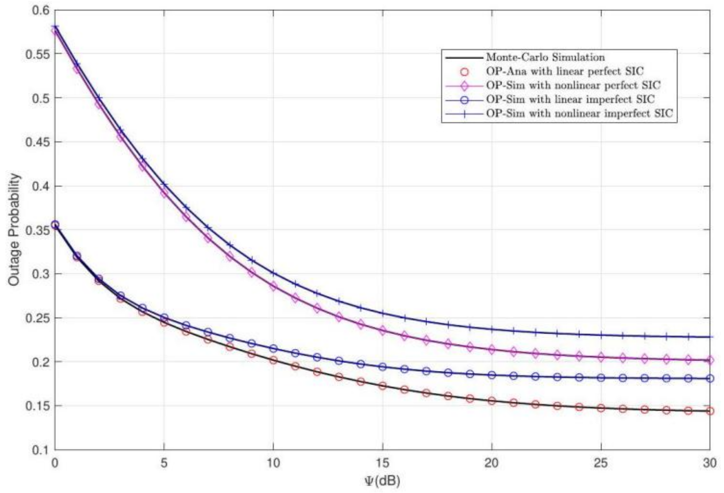


Figure 3. OP *versus* $\Psi(\text{dB})$ for linear and nonlinear models under perfect and imperfect SIC scenarios.

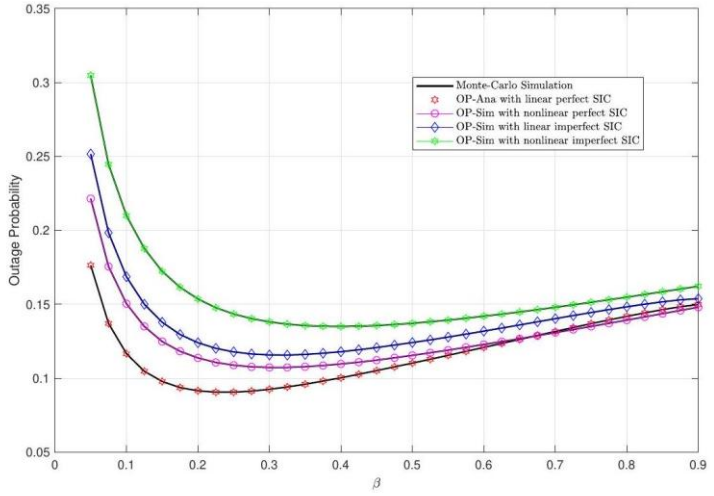


Figure 4. OP *versus* β for linear and nonlinear models under perfect and imperfect SIC scenarios.

Figure 4 illustrates the variation of the OP under the influence of the reflection coefficient β. The most prominent feature observed is the convex nature of the OP curves, which indicates the existence of a unique optimal value, denoted as β^*, that minimizes the system's outage probability. This behavior illustrates the fundamental trade-off between the EH process and the backscatter-communication efficiency. In the low β regime, a significant portion of the source power is allocated to EH, ensuring the activation of the backscatter device. However, the power fraction reserved for signal reflection is insufficient, resulting in a low SINR at the destination, thereby increasing the outage probability. Conversely, in the high β regime, while the reflected signal power is enhanced, the harvested energy at the backscatter device decreases. If the harvested power falls below the activation threshold ($P_h < P_{\text{th}}$), the circuit fails to operate, leading to system outage. From the adopted non-linear model as shown in Equation (19), the optimal value $\beta_{\text{h}}^* = \max\left(1 - \frac{\zeta}{\gamma_{SB}}, 0\right)$ is derived to ensure the successful transmission condition $P_{\text{nEH}} \geq P_{\text{th}}$. Furthermore, the comparison between different scenarios, such as imperfect SIC *versus* perfect SIC or linear *versus* non-linear models, yields results consistent with the previous findings in Figure 3. Specifically, the linear model and the perfect SIC scenario consistently establish an ideal performance limit with the lowest OP levels and *vice versa*.

Figure 5 illustrates the OP as a function of the channel parameters $\lambda_{\text{SB}} = \lambda_{\text{BD}}$ across different power-splitting factors (β). A notable observation is the non-monotonic behavior of the OP curves, which suggests that the system performance is governed by a trade-off inherent to the SIC decoding mechanism. Specifically, in the strong channel regime (low λ values), the backscattered signal acts as significant interference to the direct link decoding; consequently, a smaller splitting factor (e.g., $= 0.15$) is preferable to mitigate this interference and lower the outage probability. Conversely, in the weak channel regime (high λ values), the system becomes power-limited, where the attenuation of the backscatter link compromises decodability. In this scenario, a larger splitting factor (e.g.,$= 0.5$) is required to boost the reflected-signal strength. The intersection of these performance curves demonstrates that there is no single optimal β for all channel conditions, highlighting the necessity for channel-adaptive power-splitting strategies to minimize system outage.

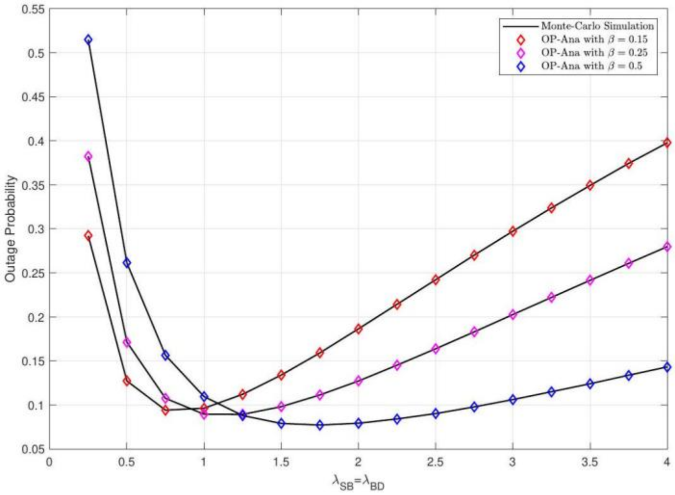


Figure 5. OP *versus* $\lambda_{\mathrm{SB}}, \lambda_{\mathrm{BD}}$, with varying β.

Figure 6 illustrates the OP as a function of the mean value of the random variable $\lambda(\lambda_{\mathrm{SB}} = \lambda_{\mathrm{BD}})$, representing the reciprocal of the mean channel gains for Rayleigh fading links. Given that the average power gain is defined as $\mathbb{E}[|\mathrm{h}|^2] = 1/\lambda$, values of $\lambda < 1$ signify favorable channel conditions with high average gains, whereas $\lambda > 1$ represents weak channel regimes or severe path loss scenarios. The graph reveals a critical performance trade-off: in the $\lambda < 1$ region, the OP is initially high, but drops rapidly as λ approaches 1, as the system moves away from being interference-limited, which typically hinders the SIC process. Conversely, as λ increases beyond the optimal point (around ≈ 1.5), the system becomes power-limited. These results are entirely consistent with the findings observed in Figure 3 and Figure 4, where the non-linear model, characterized by hardware parameters $\epsilon_0, \epsilon_1, \epsilon_2$ and the saturation power $P_{\mathrm{B}}^{\max}$, consistently yields a higher OP compared to the ideal linear model.

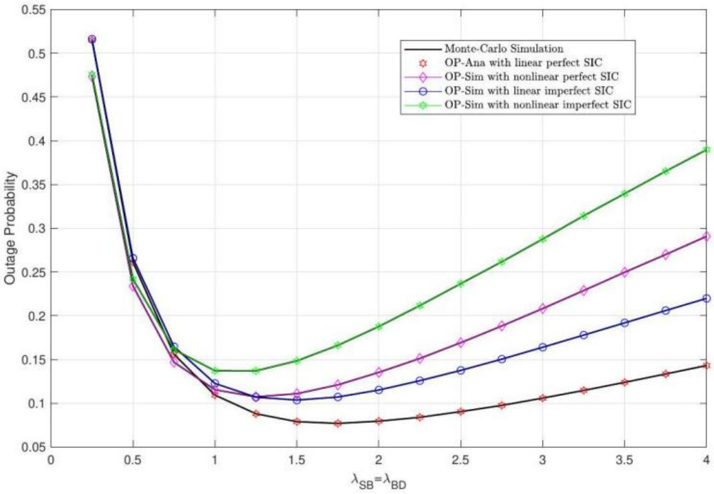


Figure 6. OP *versus* $\lambda_{\mathrm{SB}} = \lambda_{\mathrm{BD}}$ for linear and nonlinear models under perfect and imperfect SIC scenarios.

Figure 7 investigates the influence of the direct-link quality, denoted by λ_{SD}, on the system's OP under varying backscatter channel conditions ($\lambda_{\mathrm{SB}}, \lambda_{\mathrm{BD}}$). A general trend observed is the linear degradation of system performance (increasing OP) as λ_{SD} increases, corresponding to severe attenuation in the direct link from the source to the destination. However, a counter-intuitive phenomenon is evident in the relative performance of the scenarios: the configuration with the strongest backscatter channels ($\lambda_{\mathrm{SB}} = \lambda_{\mathrm{BD}} = 0.5$, red curve) yields the highest outage probability, whereas the weaker backscatter channel configuration ($\lambda_{\mathrm{SB}} = \lambda_{\mathrm{BD}} = 1.5$, blue curve) results in the best performance. This observation substantiates the interference-limited nature of the SIC process; specifically, when the direct link is weak, a strong reflected signal from the backscatter device acts as severe interference, drastically reducing the SINR required to decode the source signal x_{S}. Consequently, in scenarios with poor direct-link connectivity, mitigating interference from the backscatter branch takes precedence over maximizing reflected power, thereby significantly enhancing the overall system reliability.

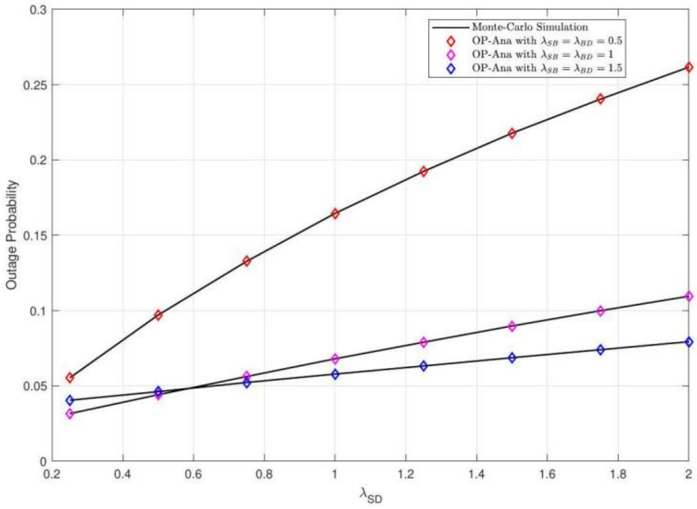


Figure 7. OP *versus* λ_{SD}, with varying $\lambda_{\mathrm{SB}}, \lambda_{\mathrm{BD}}$.

The OP as a function of the parameter λ_{SD}, which is the reciprocal of the mean channel gain for the direct source-to-destination link, is shown in Figure 8. In perfect-SIC scenarios, the OP increases steadily as λ_{SD} rises from 0.25 to 2, indicating that overall system performance decreases as the direct-channel quality deteriorates due to the weakened reliability of the direct-transmission path. The results show different system behaviors depending on interference-cancellation proficiency. Conversely, imperfect SIC scenarios exhibit an optimal point (minimum OP) around $\lambda_{\mathrm{SD}} \approx 0.75$ to 1.0. In the low λ_{SD} regime, despite high channel gains, the OP rises, because the system becomes interference-limited, exceeding the capabilities of imperfect interference cancellation; meanwhile, beyond the optimal threshold, the OP increases again as the system enters the power-limited regime. Consistent with Figures 3, 4, and 6, the non-linear model consistently exhibits higher OP than the ideal linear model due to physical constraints, such as power saturation $P_{\mathrm{B}}^{\max}$ and activation thresholds of the energy-harvesting circuit.

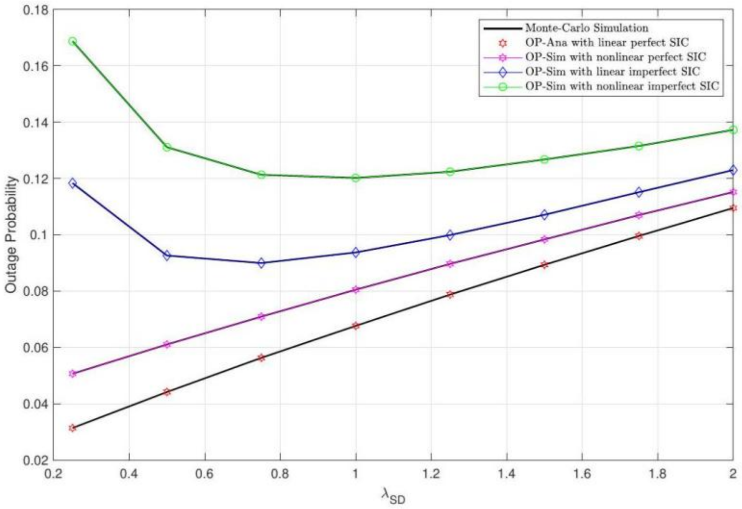


Figure 8. OP *versus* λ_{SD} for linear and nonlinear models under perfect and imperfect-SIC scenarios.

Figure 9 illustrates the variation of the OP as a function of the parameter Ψ (in dB), representing the transmit power or SNR of the system. The results show that as Ψ increases from 0 to 30 dB , the outage probability for all scenarios decreases significantly, demonstrating that increasing the transmit power is a direct solution for enhancing the reliability of the transmission link, which is consistent with the findings in Figures 2 and 3. Furthermore, the linear model consistently achieves a much lower OP compared to the corresponding nonlinear model due to physical limits imposed by hardware parameters $\epsilon_0, \epsilon_1, \epsilon_2$, similar to previous findings. Notably, emphasizing the impact of CSI, the gap between perfect-CSI and imperfect-CSI scenarios indicates that the system is highly sensitive to the accuracy of channel estimation. In practice, it is difficult to obtain perfect CSI due to channel-estimation errors (CEEs); therefore, we adopt the linear minimum mean square error method as suggested in [63] to obtain the CSI. Consequently, the communication channels are modeled following the framework in [64] as: $\hat{h}_i = h_i + e_i$ where $i \in \{\mathrm{SD}, \mathrm{BD}, \mathrm{SB}\}$ denotes the communication links, $\hat{h}_i$ is the estimated channel, and $e_i \sim \mathcal{CN}(0, \mu_i)$ represents the CEE with error variance μ_i. As Ψ approaches 30 dB, an error floor begins to emerge, where further power increments no longer yield significant reductions in OP due to the dominance of residual interference, channel-estimation inaccuracies, and nonlinear components.

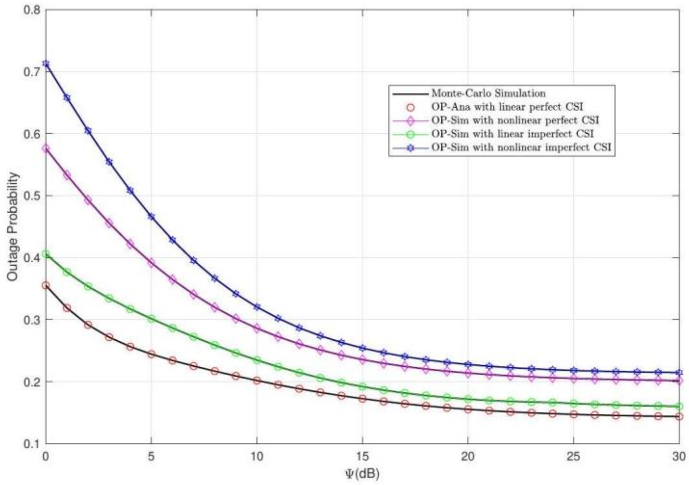


Figure 9. OP *versus* $\Psi(\mathrm{dB})$ for linear and nonlinear models under perfect and imperfect-CSI scenarios.

Table 3 summarizes the key system parameters and their impacts on the OP, highlighting the underlying physical insights revealed by the analytical and numerical results. This table provides an intuitive overview of how transmit power, target rate, power splitting, channel conditions, energy-harvesting modeling, and SIC/CSI quality jointly influence system performance.

Table 3. Summary of key insights and parameter impacts on system performance.

Parameter	**Trend**	**Key Physical Insight**
Transmit SNR (Ψ)	↘ OP	Higher power facilitates energy harvesting and improves SINR, but leads to an error floor at high values due to residual interference (Figs. 2, 3, 9).
Target Rate (R)	↗ OP	Increasing R raises the decoding threshold $\gamma_{\mathrm{th}} = 2^{\mathrm{R}} - 1$, requiring higher channel quality for success (Fig. 2).
Splitting Factor (β)	Convex	Represents the trade-off between energy harvesting (for activation) and reflection (for signal quality) (Figs. 4).
Channel Gains (λ)	Non-monotonic	System is interference-limited in strong channels ($\lambda < 1$) and power-limited in weak channels ($\lambda > 1.5$) (Figs. 5, 7, 6, 8).
EH Modeling	Value: OP of (linear < non-linear)	Linear model acts as an ideal upper bound, while non-linear hardware (saturation/thresholds) degrades performance (Figs. 3, 4, 6, 8, 9).
SIC/CSI Quality	Value: OP of (perfect < imperfect)	Residual interference and estimation errors significantly increase the outage probability in realistic scenarios (Figs. 3, 4, 6, 8, 9).

5. CONCLUSIONS

This paper has presented a comprehensive performance analysis of an ambient backscatter-assisted passive relaying system. We successfully derived accurate closed-form expressions for the OP over Rayleigh fading channels and validated the theoretical framework through extensive Monte-Carlo simulations. The study rigorously examined the influence of general system parameters, the most significant findings reveal how the interplay between non-linear energy harvesting and imperfect SIC/CSI dictates the system's reliability limits. Specifically, the results demonstrated a clear convex relationship between the OP and the power-splitting factor, confirming the existence of a unique optimal value that balances energy harvesting and signal reflection. Furthermore, our analysis underlines the detrimental impact of residual interference, which leads to the emergence of inevitable error floors in the high-SNR regime. A non-monotonic performance trend was also observed regarding the backscatter-link quality, revealing an optimal operating regime where the system effectively balances signal strength against the interference floor. These insights provide a robust quantitative tool for the design of ultra-energy-efficient IoT networks. For future work, we aim to extend this analytical framework to more complex multi-node scenarios and explore advanced beamforming techniques to further mitigate the identified performance bottlenecks.

REFERENCES

[1] S. Dang, O. Amin, B. Shihada and M.-S. Alouini, "What Should 6G Be?," Nature Electronics, vol. 3, no. 1, pp. 20-29, DOI: 10.1038/s41928-019-0355-6, Jan. 2020.
[2] H. Pennanen et al., "6G: The Intelligent Network of Everything," IEEE Access, vol. 13, pp. 1319-1421, DOI: 10.1109/ACCESS.2024.3521579, 2025.
[3] W. Jiang, B. Han, M. A. Habibi and H. D. Schotten, "The Road towards 6G: A Comprehensive Survey," IEEE Open Journal of the Communications Society, vol. 2, pp. 334-366, 2021.
[4] W. Wu et al., "A Survey on Ambient Backscatter Communications: Principles, Systems, Applications and Challenges," Computer Networks, vol. 216, p. 109235, DOI: 10.1016/j.comnet.2022.109235, 2022.
[5] M. A. Jamshed et al., "Artificial Intelligence, Ambient Backscatter Communication and Non-terrestrial Networks: A 6G Commixture," IEEE Internet of Things Magazine, vol. 8, no. 2, pp. 88-94, 2025.
[6] C. Liaskos et al., "Realizing Ambient Backscatter Communications with Intelligent Surfaces in 6G Wireless Systems," IEEE Wireless Communications, vol. 29, no. 1, pp. 178-185, Feb. 2022.
[7] H.-N. Nguyen et al., "Secure Performance Analysis of Satellite-terrestrial Networks-assisted Backscatter Device," Jordanian J. of Computers and Inform. Techn. (JJCIT), vol. 11, no. 4, pp. 484-498, Dec. 2025.
[8] D. L. Galappaththige, F. Rezaei, C. Tellambura and S. Herath, "RIS-empowered Ambient Backscatter Communication Systems," IEEE Wireless Communications Letters, vol. 12, no. 1, pp. 173-177, 2023.
[9] H. Yang et al., "A RIS-segmented Symbiotic Ambient Backscatter Communication System," IEEE Transactions on Vehicular Technology, vol. 73, no. 1, pp. 812-825, 2024.
[10] A.-T. Le et al., "Performance Analysis of RIS-assisted Ambient Backscatter Communication Systems," IEEE Wireless Communications Letters, vol. 13, no. 3, pp. 791-795, 2024.
[11] S. Jia et al., "Secrecy Performance Analysis of UAV-assisted Ambient Backscatter Communications with Jamming," IEEE Transactions on Wireless Communications, vol. 23, no. 12, pp. 18111-18125, 2024.
[12] J. Liu et al., "Intelligent Reflecting Surface-aided Covert Ambient Backscatter Communication," IEEE Trans. on Communications, vol. 72, no. 6, pp. 3558-3571, 2024.
[13] T. N. Nguyen et al., "On the Performance of Secured Ambient Backscatter Communications to Protect Digital Content and Copyrights," IEEE Access, vol. 13, pp. 195385-195400, 2025.
[14] H. Zhu et al., "Machine Learning-based Blind Signal Detection for Ambient Backscatter Communication Systems," IEEE Trans. on Cognitive Comm. and Netw., vol. 11, no. 2, pp. 1172-1183, 2025.
[15] J. Chen, Q. Guan, Y. Rong and H. Yu, "Detections for Ambient Backscatter Communications Systems with Dynamic Sources," IEEE Transactions on Communications, vol. 73, no. 9, pp. 7941-7951, 2025.
[16] J. Liao, T. Zhang, K. Ruttik, R. Jäntti and D.-T. Phan-Huy, "Ambient Backscatter Communication in LTE Uplink Sounding Reference Signal," arXiv preprint, DOI: 10.48550/arXiv.2501.10952, 2025.
[17] M. Tran et al., "Outage Analysis of a Hybrid Relay-Backscatter Communication System with Energy Harvesting for IoT and 6G Networks," IEEE Access, vol. 13, pp. 188605-188617, 2025.
[18] X. Liu, H. Wang, K. Zheng and K. Chi, "Throughput Maximization of IoT Transmission in STAR-RIS-aided Symbiotic Radio Networks," IEEE Internet of Things J., vol. 13, no. 2, pp. 3371-3388, Jan. 2026.
[19] L. S. Phu et al., "Enhancing Short-packet Communications: BLER Performance in RIS-assisted Ambient Backscatter NOMA Systems," PLOS ONE, vol. 20, no. 8, pp. 1-26, 2025.
[20] Y. Ye, L. Shi, X. Chu, G. Lu and S. Sun, "Mutualistic Cooperative Ambient Backscatter Communications under Hardware Impairments," IEEE Trans. on Communications, vol. 70, no. 11, pp. 7656-7668, 2022.
[21] K. Zheng et al., "DDPG-based Joint Time and Energy Management in Ambient Backscatter-assisted

Hybrid Underlay CRNs," IEEE Trans. on Communications, vol. 71, no. 1, pp. 441-456, Jan. 2023.
[22] T.-H. T. Pham et al., "Performance Analysis in D2D Partial NOMA-assisted Backscatter Communication," Advances in Electrical and Electronic Engineering, vol. 23, no. 3, pp. 250-262, 2025.
[23] Q.-S. Nguyen et al., "Power Beacon-assisted Energy Harvesting in D2D Network under Co-channel Interferences: Symbol Error Rate Analysis," Jordanian Journal of Computers and Information Technology (JJCIT), vol. 11, no. 4, pp. 517-532, 2025.
[24] X. Liu, J. Xu, K. Zheng, G. Zhang, J. Liu and N. Shiratori, "Throughput Maximization with an AoI Constraint in Energy Harvesting D2D-enabled Cellular Networks: An MSRA-TD3 Approach," IEEE Trans. on Wireless Communications, vol. 24, no. 2, pp. 1448-1466, Feb. 2025.
[25] G. Moloudian et al., "RF Energy Harvesting Techniques for Battery-less Wireless Sensing, Industry 4.0, and Internet of Things: A Review," IEEE Sensors Journal, vol. 24, no. 5, pp. 5732-5745, 2024.
[26] B. Y. León Ávila et al., "Energy Harvesting Techniques for Wireless Sensor Networks: A Systematic Literature Review," Energy Strategy Reviews, vol. 57, p. 101617, 2025.
[27] J. Zhou et al., "Metamaterials and Metasurfaces for Wireless Power Transfer and Energy Harvesting," Proceedings of the IEEE, vol. 110, no. 1, pp. 31-55, DOI: 10.1109/JPROC.2021.3127493, 2022.
[28] Y. Albaihani et al., "Optimal Antenna Design for Wireless Energy Harvesting System in ISM Band," Results in Physics, vol. 73, p. 108255, DOI: 10.1016/j.rinp.2025.108255, 2025.
[29] P. Zhang, X. Zhang and L. Li, "An Optically Transparent Metantenna for RF Wireless Energy Harvesting," IEEE Transactions on Antennas and Propagation, vol. 70, no. 4, pp. 2550-2560, 2022.
[30] Q.-S. Nguyen, C.-H. Tran, T.-D. Tran, M. Tran and B.-S. Kim, "Securing Wireless Communications with Energy Harvesting and Multi-antenna Diversity," Jordanian Journal of Computers and Information Technology (JJCIT), vol. 11, no. 2, pp. 197-210, DOI: 10.5455/jjcit.71-1732244909, Jun. 2025.
[31] B. V. Minh et al., "Self-energy Recycling in DF Full-duplex Relay Network: Security-Reliability Analysis," Advances in Electrical and Electronic Engineering, vol. 22, no. 1, pp. 86-96, 2024.
[32] S. Ghosh et al., "On the Performance of End-to-end Cooperative NOMA-based IoT Networks with Wireless Energy Harvesting," IEEE Internet of Things Journal, vol. 10, no. 18, pp. 16253-16270, 2023.
[33] D. Bepari et al., "Uplink Performance Analysis of Wireless Energy Harvesting-enabled NOMA-based Networks," Mobile Networks and Applications, vol. 29, no. 3, pp. 856-866, 2024.
[34] T. N. Nguyen et al., "On the Dilemma of Reliability or Security in Unmanned Aerial Vehicle Communications Assisted by Energy Harvesting Relaying," IEEE J. on Selected Areas in Comm., vol. 42, no. 1, pp. 52-67, 2024.
[35] X. Liu et al., "AoI-minimal Clustering, Transmission and Trajectory Co-design for UAV-assisted WPCNs," IEEE Trans. on Vehicular Technology, vol. 74, no. 1, pp. 1035-1051, Jan. 2025.
[36] K. Ntontin et al., "Wireless Energy Harvesting for Autonomous Reconfigurable Intelligent Surfaces," IEEE Trans. on Green Communications and Networking, vol. 7, no. 1, pp. 114-129, 2023.
[37] M. Poposka et al., "Design Optimization of RF Energy Harvesting Networks for Federated Learning," Proc. 2024 Int. Balkan Conf. Commun. Networking (BalkanCom), pp. 58-62, DOI: 10.1109/BalkanCom61808.2024.10557202, 2024.
[38] B. V. Minh, N. H. K. Nhan, T.-H. T. Pham, M. Tran and S.-W. Kim, "Physical Layer Security in Wireless Sensors Networks with Friendly Jammer: Secrecy Outage Probability Analysis," Advances in Electrical and Electronic Engineering, vol. 22, no. 4, pp. 387-398, DOI: 10.15598/aeee.v22i4.5840, 2024.
[39] M. Malik, A. Kothari and R. Pandhare, "Smart Military Logistics Based on Internet of Things and Energy Harvesting," Advances in Electrical & Electronic Engineering, vol. 23, no. 2, 2025.
[40] R. Du et al., "Comparing Backscatter Communication and Energy Harvesting in Massive IoT Networks," IEEE Transactions on Wireless Communications, vol. 21, no. 1, pp. 429-443, 2022.
[41] J. Zan et al., "Stochastic Geometry Based Performance Study for Wireless Powered Backscatter Communications," IEEE Trans. on Vehicular Technology, vol. 71, no. 10, pp. 11136-11149, 2022.
[42] T. Jiang et al., "Backscatter Communication Meets Practical Battery-free Internet of Things: A Survey and Outlook," IEEE Communications Surveys & Tutorials, vol. 25, no. 3, pp. 2021-2051, 2023.
[43] H. Ma et al., "Reconfigurable Intelligent Surface with Energy Harvesting Assisted Cooperative Ambient Backscatter Communications," IEEE Wireless Comm. Letters, vol. 11, no. 6, pp. 1283-1287, 2022.
[44] X. Liu, X. Li, K. Zheng and J. Liu, "AoI Minimization of Ambient Backscatter-assisted EHCRN with Cooperative Spectrum Sensing," Computer Networks, vol. 245, p. 110389, 2024.
[45] K. Zheng et al., "A Hybrid Communication Scheme for Throughput Maximization in Backscatter-aided Energy Harvesting Cognitive Radio Networks," IEEE IoT J., vol. 10, no. 18, pp. 16194-16208, 2023.
[46] X. Liu etal., "Optimal Time Allocation for Backscatter-aided Relay Cooperative Transmission in Wireless-powered Heterogeneous CRNs," IEEE IoT J., vol. 10, no. 18, pp. 16209-16224, 2023.
[47] A. Iqbal and T.-J. Lee, "Opportunistic Backscatter Communication Protocol Underlying Energy Harvesting IoT Networks," IEEE Access, vol. 11, pp. 89568-89580, 2023.
[48] W.-J. Wang, K. Xu, Y. Yan and L. Chen, "Relay Selection-based Cooperative Backscatter Transmission with Energy Harvesting: Throughput Maximization," IEEE Wireless Communications Letters, vol. 11, no. 7, pp. 1533-1537, DOI: 10.1109/LWC.2022.3179019, 2022.

[49] X. Liu et al., "Throughput Maximization of Wireless-powered Communication Network with Mobile Access Points," IEEE Transactions on Wireless Communications, vol. 22, no. 7, pp. 4401-4415, 2023.
[50] B. Lyu, C. You, Z. Yang and G. Gui, "The Optimal Control Policy for RF-powered Backscatter Communication Networks," IEEE Trans. on Vehicular Technology, vol. 67, no. 3, pp. 2804-2808, 2018.
[51] P. Ghosh, H. Yenna, S. D. Roy and S. Kundu, "Outage Analysis of an EH Relay Aided Network with Ambient-Backscattering," Proc. IEEE Int. Conf. Electronics, Computing and Communication Technologies (CONECCT), pp. 1-6, DOI: 10.1109/CONECCT62155.2024.10677243, 2024.
[52] L. Shi, J. Shi, Y. Ye, G. Zheng and G. Lu, "Ambient Backscatter Communication with HARQ Assisted Hybrid Long-short Packets," IEEE Communications Letters, vol. 28, no. 10, pp. 2258-2262, 2024.
[53] J. Shi et al., "Performance Analysis for Ambient Backscatter Communications with Hybrid Long-short Packets," IEEE Wireless Communications Letters, vol. 13, no. 5, pp. 1325-1329, 2024.
[54] X. Song et al., "Relay Assisted Cooperative Ambient Backscatter Communication with Hybrid Long-short Packets," IEEE Trans. on Vehicular Technology, vol. 73, no. 9, pp. 12890-12903, 2024.
[55] P. Ghosh, S. D. Roy and S. Kundu, "Energy Harvesting-assisted Two-user Cooperative NOMA with Ambient Backscattering," Int. J. of Communication Systems, vol. 38, no. 4, p. e6133, 2025.
[56] B. C. Nguyen et al., "Improving the Performance of Spatial Modulation Fullduplex Relaying System with Hardware Impairment Using Transmit Antenna Selection," IEEE Access, vol. 8, pp. 20191-20202, 2020.
[57] M. H. Tran, B. C. Nguyen and T. T. Phuong, "Outage Analysis of RF Energy Harvesting Cooperative Communication Systems over Nakagami-fading Channels with Integer and Non-Integer m," IEEE Transactions on Vehicular Technology, vol. 69, no. 3, pp. 2785-2801, Jan. 2020.
[58] T. N. Nguyen, T. T. Phuong and M. Voznak, "Wireless Energy Harvesting Meets Receiver Diversity: A Successful Approach for Two-way Half-duplex Relay Networks over Block Rayleigh Fading Channel," Computer Networks, vol. 172, p. 107176, DOI: 10.1016/j.comnet.2020.107176, May 2020.
[59] H. Nguyen et al., "Security-Reliability Analysis in CR-NOMA IoT Network under I/Q Imbalance," IEEE Access, vol. 11, pp. 119045-119056, DOI: 10.1109/ACCESS.2023.3327789, Nov. 2023.
[60] N.-T. Nguyen et al., "Performance Analysis of NOMA-based Hybrid Satellite-Terrestrial Relay System Using mmWave Technology," IEEE Access, vol. 11, pp. 10696-10707, Jan. 2023.
[61] Y. Khan et al., "Secrecy Analysis of Energy Harvesting Backscatter Communication Networks with Multiple Eavesdroppers and Different Tag Selection Schemes," IEEE Transactions on Green Communications and Networking, pp. 1-15, DOI: 10.1109/TGCN.2025.3563107, 2025.
[62] T. T. Duy et al., "On the Performance of the Coverage Probability of LoRa Networks with Non-linear Energy Harvesting," Proc. 2024 Int. Conf. on Advanced Technologies for Communications (ATC), pp. 722-726, DOI: 10.1109/ATC63255.2024.10908157, Ho Chi Minh City, Vietnam, 2024.
[63] X. Li et al., "Security and Reliability Performance Analysis of Cooperative Multi-relay Systems with Nonlinear Energy Harvesters and Hardware Impairments," IEEE Access, vol. 7, pp. 102644-102661, DOI: 10.1109/ACCESS.2019.2930664, 2019.
[64] T. N. Nguyen et al., "Security and Reliability Analysis of Satellite-terrestrial Multi-relay Networks with Imperfect CSI," IEEE Systems Journal, vol. 17, no. 2, pp. 2824-2835, Aug. 2022.

ملخص البحث:

تتناول هذه الورقة بِنية مبتكرة تستخدم بروتوكولات حصاد الطّاقة لدمْج تقنية الاتّصال بالتّشتُّت الخلفي المحيط كجهاز ترحيلٍ سلبي. وهذا الجهاز ذو الطّاقة المحدودة يستخدم آليةً لتقسيم الطّاقة لتفعيل الدّارة. ويتمثل الاسهام لهذا العمل في تطوير صيغ رياضية مغلقة جديدة ودقيقة لاحتمالية انقطاع النّظام عبر قنوات (رايلي) المتلاشّية. وأجريت محاكاة (مونتي كارلو) مكثّفة للتّحقُّق بدقّةٍ من صحّة الإطار التّحليلي المقترح. وكشف التّحليل مُفاضلاتٍ مهمّة بين موثوقية الإرسال وكفاءة حصاد الطّاقة، الأمر الّذي يوفّر رؤى قيمّة لتحسين استخدام الموارد في شبكات إنترنت الأشياء منخفضة الطّاقة المستقبلية. وأظهرت النّتائج إمكانية التّخفيف الفعّال من الآثار السّلبية لعدم اكتمال إلغاء التّداخل المتتالي و/أو عدم اكتمال معلومات حالة القناة عن طريق زيادة طاقة الإرسال و/أو التّشغيل عند القيمة المثلى لمعامل الانعكاس. واتّضح أنّ فجوة الأداء بين نظامي (SIC) و(CSI) المثاليّيْن وغير المثاليّيْن صغيرة نسبياً. وأخيراً، نبرهن تحليلياً أنّ نموذج حصاد الطّاقة الخطّي يمثّل حدّاً أعلى لنموذج حصاد الطّاقة غير الخطّي العملي.

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 12, No. 01, March 2026.

BRAIN-TUMOR CLASSIFICATION USING RESNET50 ENHANCED WITH SE AND CBAM ATTENTION MECHANISMS

Nadia Shamsulddin Abdulsattar[1] and Fatimah S. Abdulsattar[2]

(Received: 5-Dec.-2025, Revised: 15-Feb.-2026, Accepted: 9-Mar.-2026)

ABSTRACT

MRI image classification of brain tumors is critical for accurate and early diagnosis. New developments in deep learning have revealed that inserting attention mechanisms into convolutional neural networks can greatly improve classification performance. The ECA attention mechanism is also introduced in this study. This work assesses the effectiveness of Squeeze-and-Excitation (SE) and Convolutional Block Attention Module (CBAM) sequentially integrated with the ResNet50 model, which increases classification accuracy, precision and recall when compared to the basic model, according to experimental results on two datasets for brain tumors. The suggested model employs attention mechanisms to focus valuable information selectively and suppress irrelevant information. The experiments are conducted on two datasets (Brain Tumor MRI and Brisc). The first dataset displays great improvements over basic CNN models, with precision, recall, accuracy, F1 score and AUC at 0.9914, 0.9903, 0.9945, 0.9908 and 0.9989, respectively. The second dataset gives the results for precision, recall, accuracy, F1 score and AUC at 0.9860, 0.9857, 0.9860, 0.9858 and 0.9985, respectively. From these results, the importance of attention mechanisms in deep-learning models for medical imaging is highlighted, which suggests that SE and CBAM modules can be available as more dependable and effective instruments for brain-tumor classification in clinical settings. Future studies should investigate transformer-based and hybrid attention techniques to enhance automated brain tumor categorization.

KEYWORDS

Brain Tumor, MRI, ResNet50, Squeeze-and-Excitation (SE), CBAM.

1. INTRODUCTION

Many studies have focused on the detection of brain tumors in recent years [1][2][3][4][5][6][7][8][9][10][11]. Researchers have suggested some techniques to reveal brain tumors in MRI scans of the brain [12]. The accuracy of each of these techniques has varied. Convolutional neural networks (CNNs), a common deep-learning technique, have an advantage over classical networks. These can learn from the input directly without needing human feature extraction [13]-[14]. The CNN model is composed of several layers, each with a distinct job. Features are estimated by the convolution layer, the size of the features from the previous layer is reduced by the pooling layer, the features of high-level are elicited and the output of the model is predicted by the fully connected layer, also known as the dense layer. The basic structure of CNN employs activation functions, like Tanh, Sigmoid and ReLU [15]-[16] and [17]. Brain tumors, like meningiomas, gliomas and pituitary adenomas, are among the brain tumors that pose serious health risks and need to be precisely identified to receive focused treatments. The clinical gold standard for non-invasive tumor visualization and evaluation is still magnetic resonance imaging (MRI). Recent advances in deep learning; namely, convolutional neural networks (CNNs), have revolutionized computer-aided diagnosis in neuroimaging by offering automated solutions that can match or surpass human expert performance when given sufficient, well-annotated data [18]. Optimizing feature extraction from intricate MRI data is a great challenge in the automated classification of brain tumors. Although standard CNNs function well, they can be improved by using attention techniques that direct the network to more noticeable aspects of the image [19]. Several studies have explored CNN architectures for brain-tumor classification. ResNet variants are widely employed due to their depth and performance. SE networks improve channel interdependencies by recalibrating feature maps. CBAM combines channel and spatial attention to further improve discriminative feature

---

1. N. S. Abdulsattar is with Department of Computer, College of Basic Education, Mustansiriyah University, Baghdad, Iraq. Email: nadiashamsaldeen@uomustansiriyah.edu.iq
2. F. S. Abdulsattar is with Department of Computer Engineering, College of Engineering, Mustansiriyah University, Baghdad, Iraq. Email: fsa.abdulsattar@uomustansiriyah.edu.iq

representation. To enhance the representational capacity of the baseline ResNet50 model, channel and spatial-attention mechanisms were incorporated into the residual blocks. Sequential integration of SE and CBAM was used to exploit their complementary channel and spatial-attention mechanisms while avoiding attention redundancy and improving model generalization.

2. RELATED WORK

A method for detecting and categorizing brain tumors using Deep Residual Networks (ResNet) was presented by Sahaai et al. [20]. In this method, a ResNet50 model based on transfer learning and the CNN architecture is employed to achieve multi-class classification of brain tumors. Through several brain-tumor dataset categories, this method obtains 95.3%, 94.6%, 92.2%, 93.7% and 87.8% for accuracy, F1 score, recall, precision and specificity, respectively.

Oladimeji and Ibitoye [21] employed the pretrained model ResNet50 with the CBAM. When comparing this method with other classification methods that use the same dataset (Brain Tumors dataset), the ResNet50-CBAM achieved good results compared with existing deep-learning classification methods, such as CNNs. It performed a recall, accuracy, precision and AUC at 99.01%, 99.43%, 98.7% and 99.25%, respectively. The study's experimental findings demonstrated the convolutional block attention mechanism scheme's better results in brain-tumor categorization.

To categorize brain tumors, Vinston et al. [22] employed the pretrained ResNet50 with the CBAM. This method outperformed previous techniques, like traditional convolutional neural networks. On the same dataset (Brain Tumors dataset), the ResNet50-CBAM model revealed good results, reaching 99.53%, 99.11%, 99.35% and 98.75% for area under the curve (AUC), recall, accuracy and precision, respectively. The convolutional block attention-mechanism architecture works exceptionally well for brain-tumor classification, according to experimental results.

Employing the dataset on Kaggle, which contains 3,096 MRI images. Huang and Prakash [23] trained a ResNet50V2-based model. They evaluated five models: Basic ResNet50V2, Squeeze-and-Excitation, Convolutional Block Attention Module, Self-Attention and Attention Gated Network. To compare the classification accuracy of the models, two proportion Z-tests were employed. The SE model surpassed the basic ResNet50V2 in classification, achieving an accuracy of about 98.4% and an AUC of 1.00, while the ResNet50V2 achieved about 92.6%. In this study, SE achieved the highest results compared to the other attention mechanisms, such as CBAM.

To categorize brain tumors independently, Rakesh et al. [24] employed a complex convolutional neural network. They used a set of methods to process the dataset, like cropping, splitting and uncropping. The ResNet-50 pretrained model is the primary model employed in this method. With 81.67%, 74.3%, 82.55% and 81.67% for accuracy, precision, recall and specificity, the suggested model performs admirably and exceeds early predictions.

To categorize brain tumors, Md and Ankit [25] created and evaluated custom convolutional neural networks. Our findings showed that these tailored models beat popular structures, like ResNet18 and VGG16, in accuracy and efficiency of computation while maintaining a simpler design. The custom CNN fulfilled 98.09% accuracy in multi-class classification, 98.67% on the Br35H dataset and 99.62% on the Brain Tumor MRI dataset in binary classification. The custom CNNs offered an additional computationally efficient option, while ResNet18 and VGG16 kept good performance levels in comparison.

Muhammad et al. [26] suggested CNNs, which are excellent at extracting features at the convolutional layer. The architectures of the outstanding CNNs employed in medical image processing, such as VGG16, ResNet-50, MobileNet, InceptionV3 and EfficientNetB7, are compared with the brain-tumor classification job. With the result, VGG16 gives good findings on other CNN architectures. For test-set data, VGG16 produces 100% accuracy, sensitivity, specificity, precision and F1-score. This study indicates the effectiveness of the strategy suggested by showing outstanding performance in categorizing brain tumors and no tumor from MRI scans.

The objective of the study of Mohammad and Muhammet [27] was to categorize brain cancers, like gliomas, meningiomas and pituitary tumors, using the images of brain MRI. The CNNs and CNN-dependent categorization methods included transfer learning, Inception-V3, EfficientNetB4 and VGG19. F1 score, recall, accuracy and imprinting were employed to assess these models. VGG16 yielded the best results, with an accuracy rate of 98%. The same transfer-learning model has an F1 score of 97%, an area under the curve of 99%, a recall of 98% and a precision of 98%. The CNN structure and CNN-dependent transfer-learning models are essential to the health of humans for the detection and timely medication of diseases early.

In contrast to the previous three suggested models (InceptionResNet, DenseNet121 and NasNet Large), Asif et al. [28] suggested a model of CNN that depends on the architecture of the Xception that utilizes the optimizer (ADAM). Sensitivity, accuracy, precision, specificity and F1-score values for the Xception model were 99.68%, 99.67%, 99.68%, 99.66% and 99.68% on the dataset (BR35H) and 96.55%, 91.94%, 87.50%, 87.88% and 91.80% on the dataset (Brain Tumor). The suggested approach performs well compared to those in the existing literature in detecting brain tumors.

To elicit features from the images of brain MRI, Deepak and Ameer [29] employed a pre-trained GoogLeNet and adopted the idea of deep transfer learning. The obtained features are classified using integrated, validated classifier models. The experiment utilized a five-fold cross-validation approach at the level of patient, using the MRI dataset from figshare. The proposed approach achieved a good result compared with the cutting-edge techniques, achieving an accuracy of 98%. The other performance measures employed in the work stand for recall, specificity, AUC, F-score and precision. The study's findings suggested that transfer learning is beneficial in situations where medical images are rare.

The CNN technique, known as en-CNN, was proposed by Hapsari et al. [30]. This approach depends on VGG-16, which has four ReLU layers, four max pooling layers and seven convolutional layers. The three steps of the novel method are augmentation, preprocessing and the application of an en-CNN. The 4 MRI sequences: FLAIR, T1, T2 and T1CE are also employed in our suggested approach for classification. Employing the ADAM optimizer, the suggested approach achieves an accuracy of the multi-sequence MRI dataset (BraTS 2018) with mini-batch size 128 and epoch 200 of 95.5% for T1, 94% for T2, 95.5% for T1CE and 97% for FLAIR. Compared to earlier studies using the same dataset, the accuracy was 4% higher.

Lin et al. [31] enhanced the ResNet50 model by preprocessing the image and employing fractional calculus; then, transfer learning and the attention mechanism (ECA) are applied. After that, the enhanced ResNet50 is collected with EfficientNetB0 to increase the accuracy. The enhanced ResNet50 increased the accuracy, precision, recall and F1 score to 98.78%, 98.82%, 98.68% and 98.75%, respectively and the value of Kappa was raised to 4.7%.

Yan et al. [32] suggested a new model that fuses (enhanced MobileNetV1 and EfficientNetB0). To improve the ability of feature extraction, add local-global attention after the top-level feature map. It uses transfer learning in EfficientNetB0 and a (3 × 3) convolution is added to the residual shortcut. The results showed that the new model achieved 94.58% in classification, 94.64% in precision and 93.22% in the coefficient of Kappa.

Prashantha and Prakash [33] suggested a model to identify the best handcrafted features employing a genetic algorithm and a finetuned CNN using three datasets: (TWB-HM, RD and TCIA-IXI). The findings of the suggested model performed well compared with state-of-the-art methods, achieving 99.40% accuracy with the DCA method on the RD dataset.

Benbakreti et al. [34] designed experiments to diagnose masses in breast images using three pretrained models (ResNet18, InceptionV3 and AlexNet) on three merged datasets (DDSM, MIAS and Inbreast). The ResNet18 achieved good results of 95% accuracy, 94.91% recall, 94.90% precision and 94.91% F1 Score.

3. METHOD

The architecture of the suggested deep-learning approach for classifying brain tumors is covered in this part. It combines a ResNet50 model with two attention mechanisms: Squeeze-and-Excitation (SE) blocks and Convolutional Block Attention Modules (CBAM).

The Attention mechanisms substantially improve (MRI) tumor categorization. This approach increases classification accuracy by improving the model's capacity to extract significant spatial and channel-wise characteristics of MRI data. These three strong deep-learning components are integrated in an architecture called the ResNet50-SE-CBAM to significantly boost classification performance, feature extraction and model interpretability, particularly for difficult tasks, such as brain-tumor classification from MRI scans.

This work employs the Residual Network to extract features from preprocessed pictures employing pretrained ImageNet weights [35]. This model is composed of convolutional layers, identity blocks and a final softmax layer. The convolutional and max-pooling layer weights were locked for stability and to prevent changes during the experiment. ResNet has been chosen over pretrained networks because of its better performance and its ability to manage the vanishing gradient problem with skip connections. SE Blocks (Squeeze-and-Excitation) [36] are applied after residual blocks, which add channel attention: (Squeeze: Global Average Pooling, Excitation: Fully connected (FC) layers with ReLU + Sigmoid). Multiplies the original feature map by learned channel weights. That effect enhances important feature channels and suppresses irrelevant ones. CBAM (Convolutional Block Attention Module) [37] is applied after SE blocks to complement channel attention with spatial attention, consisting of two sub-modules: (Channel Attention and Spatial Attention). Channel Attention (similar to SE, but lighter) and Spatial Attention are applied to 2D convolution on concatenated average/max-pooled spatial descriptors, which produces a spatial attention map to focus on informative regions. That effect helps the model localize tumor regions in MRI scans.

In this work, the SE and CBAM attention modules were not integrated inside each residual bottleneck block, but rather were incorporated into the ResNet50 backbone at particular intermediate feature stages. The (conv3_block4_out) layer's output, which is the last block of the (Conv3_x) stage, was subjected to the SE module, yielding feature maps with dimensions of (28 x 28 x 512). In order to represent channel interdependencies and produce adaptive channel weights, the SE block now employs global average pooling for channel-wise recalibration, followed by a two-layer fully linked bottleneck structure. Before extracting deeper features, this improves mid-level semantic representations. The CBAM module was applied to the output of the (conv5_block3_out) layer, which yields high-level feature maps with dimensions of (7×7×2048). This layer is the final block of the (Conv5_x) stage. Channel attention and spatial attention are applied successively by CBAM in this position. Before shared fully connected layers, channel attention is calculated using both global average pooling and global max pooling. A convolutional layer with a (7×7) kernel is then used to produce spatial attention in order to highlight informative spatial locations. This allows the network to fine-tune high-level discriminative features before the final classification layers and Global Average Pooling. The integration method was depth-aware and sequential: CBAM was used at a deeper stage to improve both channel and spatial discrimination in high-level features, while SE was used at an intermediate stage to reinforce channel-feature learning in mid-level representations.

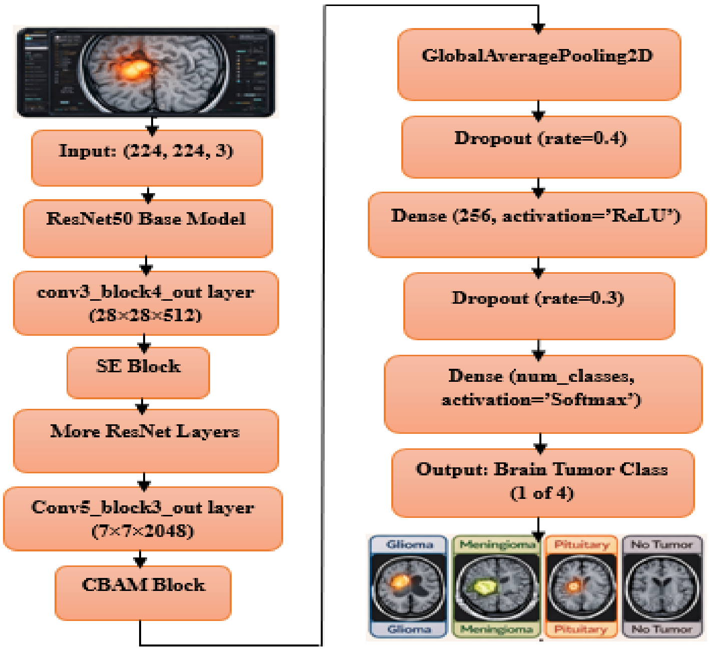


Figure 1. ResNet50-SE-CBAM architecture.

This selective stage-level integration reduces computing complexity and attention redundancy while preserving the stability of the original residual learning structure, in contrast to approaches that stack numerous attention modules within each residual block. (224 × 224) pixels make up the model's input, which is scaled with 1.0 / 255 to normalize the pixel values to the range [0, 1]. Figure 1 clearly illustrates the paradigm of the suggested approach for classifying brain tumors.

The framework of the proposed model is seen in the Figure above: A frozen, pretrained ResNet50 processes the input image. It applies the CBAM block at the deep feature layer (conv5 block3_out) and the SE block at the intermediate feature layer (conv3 block4_out). The subsequent layers are: dense, dropout, global average pooling and softmax classification. GlobalAveragePooling2D keeps spatially distributed information besides converting 3D feature maps into 1D vectors. Dropout lowers the overfitting effect. To better the decision boundaries, a dense layer incorporates learnable parameters.

4. RESULTS AND DISCUSSION

4.1 Datasets

Two datasets are employed in this work: the first dataset (Brain Tumor MRI) from Kaggle, which works with ResNet50 and other comparable CNN architectures. This collection consists of 7,023 MRI pictures of human brains classified into 4 tumor classes: “pituitary, meningioma, glioma and no tumor”. The “figshare, SARTAJ and Br35H” datasets were incorporated to make this dataset [38]. Because acquisition settings, institutions and many scanners are mixed, visual diversity is expanded, which contributes to robust model training. The dataset summary is provided in Table 1.

Table 1. Dataset summary of brain tumors.

Class	Training	Testing	Total
pituitary	1,457	300	1,757
meningioma	1,339	306	1,645
glioma	1,321	300	1,621
no tumor	1,595	405	2,000

The second dataset (Brisc) consists of 6,000 MRI pictures, also classified into 4 tumor classes: “pituitary, meningioma, glioma and no tumor”. This set consists of super-quality data that has been expertly annotated for classifying and segmenting the brain tumors. It addresses typical issues in current datasets (Figshare, BraTS), such as inconsistent annotation, narrow tumor focus and class imbalance. A realistic assessment of model generalization to unknown clinical data is made feasible by this dataset's entirely distinct distribution of brain MRI pictures that have never been seen in training sources [39]. The dataset summary is provided in Table 2.

Table 2. Dataset summary of Brisc.

Class	Training	Testing	Total
pituitary	1,457	300	1,757
meningioma	1,329	306	1,635
glioma	1,147	254	1,401
no tumor	1,067	140	1,207

This distribution is obviously skewed when compared with the initial dataset; in particular, the fraction of images that contain no tumor significantly decreased. This displays an additional obstacle for evaluating the reliability and resilience of classification models when real-world class imbalance occurs.

It applies ImageNet pre-processing for normalization and resizes each image to (224, 224) to meet the ResNet50 input specifications. During the training stage, it makes use of data-augmentation methods comprising rotation, flipping, translation and zooming. 20% of the dataset is employed for validation and 80% (remaining) is utilized for training. A collection of brain-image classes is demonstrated in Figure 2.

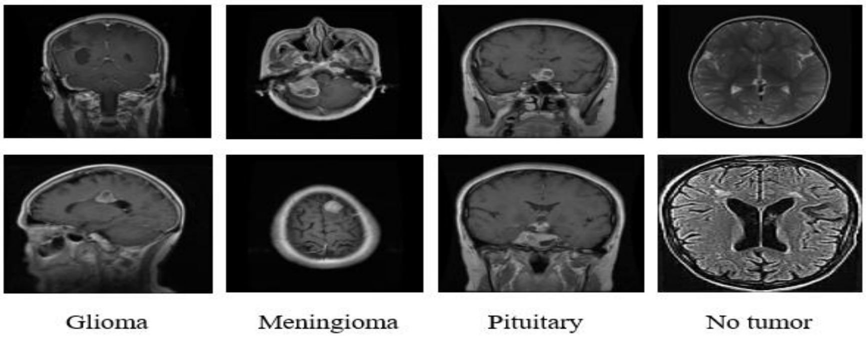


Figure 2. Several categories of brain images.

4.2 Experimental Setup

The proposed ResNet50 improved with “Squeeze-and-Excitation (SE) and Convolutional Block Attention Module (CBAM)”, was assessed on two benchmark medical imaging datasets: “a Brain Tumor MRI dataset and the BRISC dataset”. The tests employed transfer learning from pre-trained ImageNet weights and the model's upper layers were subsequently fine-tuned (the last residual block of ResNet50). A fair evaluation was ensured by dividing the dataset into a 20% validation set and an 80% training set. Generalization was enhanced by the use of data-augmentation methods, which stand for flipping, zooming and random rotations. The model's classification performance was outstanding across all evaluation metrics. The TensorFlow Python module was used to write all of the code. To run all the codes, the usual Google Colab platform with a GPU (T4) accelerator was used.

4.3 Training and Validation Performance

To train all models, we employed the AdamW optimizer. To raise stability and computational efficiency on GPU hardware, the training employed an “automatic mixed precision” method. In this work, the number of epochs and batch size were experimentally chosen depending on validation performance and convergence behaviour. Due to GPU memory limitations and its demonstrated efficacy in small-scale medical imaging datasets, a batch size of 16 was selected. The model converged within 35 epochs of training and additional training did not substantially increase validation accuracy, suggesting diminishing returns and possible overfitting. The additional parameters used during the training process are displayed in Table 3.

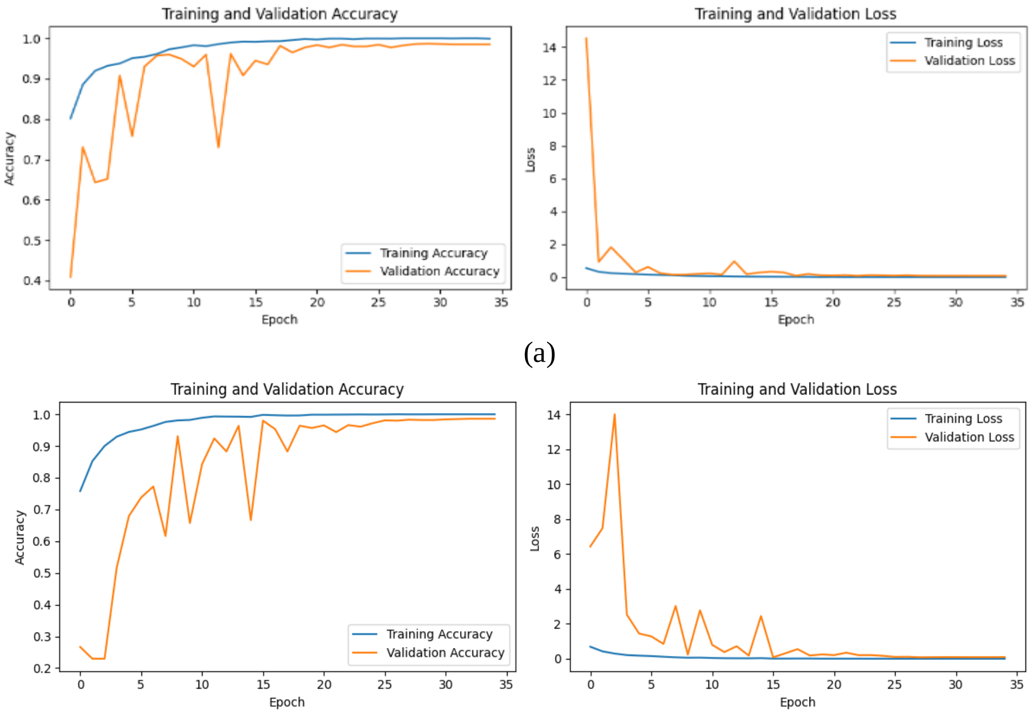
Table 3. The training parameters for the proposed model.

Parameters	**The value**
Optimizer	AdamW
Batch size	16
Epochs	35
Loss function	categorical_crossentropy
Learning rate	1e-4
Hidden-layer activation function	Relu
Classification-activation function	Softmax

All models were trained on two datasets (Brain Tumor and Brisc). The suggested model can be seen in the accuracy and loss behaviour in Figure 3.

Each model was then developed and validated using a five-fold cross-validation method. Use a 5-fold to prevent overfitting and demonstrate the model's robustness. For each fold, the metrics report the mean ± standard deviation over three runs (See Table 4). Both datasets were subjected to 5-fold cross-validation in order to assess model robustness and minimize overfitting bias. The suggested ResNet50+SE+CBAM model demonstrated solid generalization performance by achieving the highest mean accuracy with the lowest standard deviation.

Strong baselines, CNN models, such as InceptionV3 and DenseNet121, were included to contextualize the gains (See Table 5). This demonstrates that the suggested model is competitive, even when compared to CNN baselines.



(a)

(b)

Figure 3. The proposed model's changing performance over time in terms of accuracy and loss on two datasets. (a) Brain Tumor dataset. (b) Brisc dataset.

Table 4. Cross-validation performance for each model on two datasets.

Dataset	Model	K-Fold	Accuracy (Mean ± Std)	Precision (Mean±Std)	Recall (Mean±Std)	F1-score (Mean±Std)
Brain Tumor MRI	ResNet50	5-Fold	0.9501±0.0038	0.9509±0.0150	0.9484±0.0226	0.9496±0.0087
	ResNet50+SE	5-Fold	0.9764±0.0024	0.9750±0.0098	0.9702±0.0164	0.9726±0.0103
	ResNet50+CBAM	5-Fold	0.9809±0.0019	0.9800±0.0064	0.9796±0.0077	0.9798±0.0072
	ResNet50+SE+CBAM	5-Fold	0.9945±0.0012	0.9914±0.0076	0.9903±0.0080	0.9908±0.0080
Brisc	ResNet50	5-Fold	0.9540±0.0035	0.9535±0.0085	0.9533±0.0092	0.9534±0.0092
	ResNet50+SE	5-Fold	0.9639±0.0029	0.9639±0.0081	0.9629±0.0279	0.9634±0.0106
	ResNet50+CBAM	5-Fold	0.9830±0.0018	0.9830±0.0107	0.9825±0.0165	0.9827±0.0158
	ResNet50+SE+CBAM	5-Fold	0.9860±0.0015	0.9860±0.0074	0.9857±0.0093	0.9858±0.0086

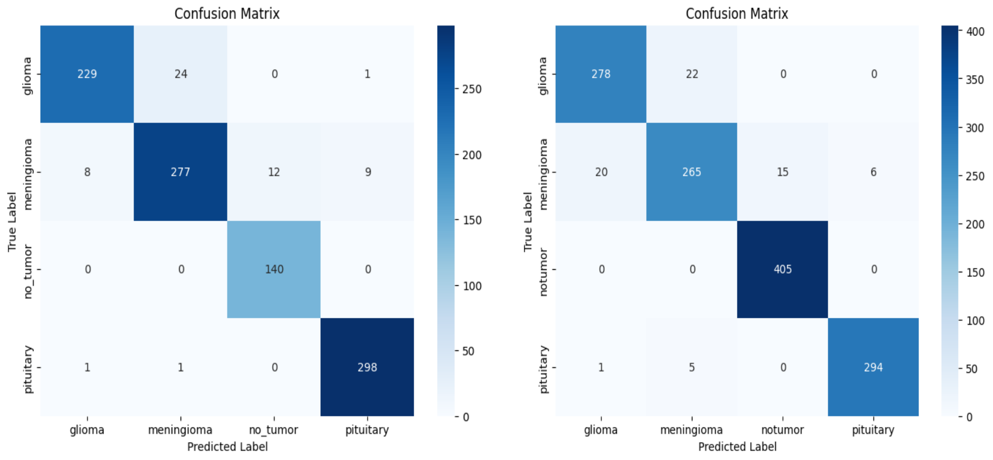
Table 5. The table of baseline comparison.

Model	MRI Accuracy	Brisc Accuracy
InceptionV3	0.9856	0.9845
DeseNet121	0.9849	0.9840
ResNet50+SE+CBAM	0.9945	0.9860

4.4 Performance Evaluation

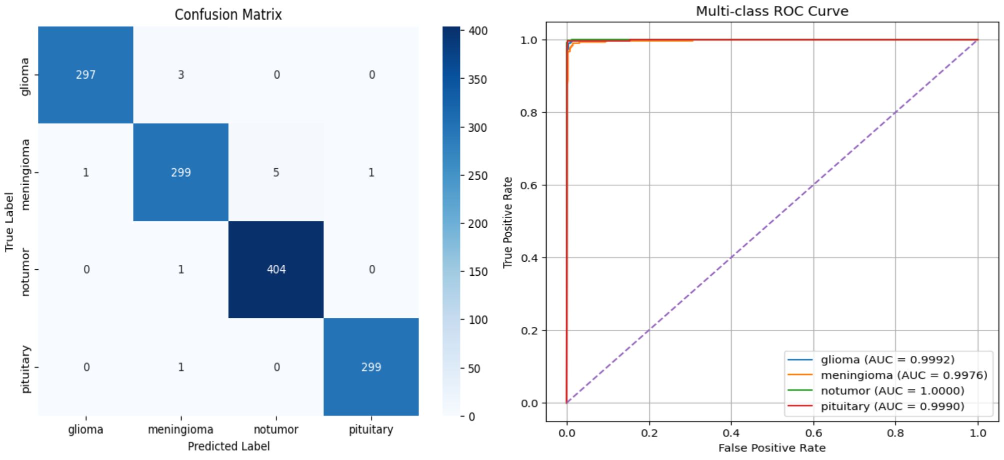
A confusion matrix and a Receiver Operating Characteristic (ROC) curve were employed for additional analysis of the classification performance. The confusion matrix exhibits how well the model can recognize various types of brain tumors, such as glioma, meningioma, pituitary and normal instances. The robustness of the suggested approach is confirmed by high “true positive and true negative” rates. Moreover, the model's great discriminative capacity across all classes is demonstrated by ROC curves and Area Under the Curve (AUC) values, underscoring its dependability in clinical-diagnosis settings.

Figure 4 shows the confusion matrix for the baseline ResNet50 model. Figures 5 and 6 show the confusion matrix and ROC curve for the proposed model. The results display a significant reduction in misclassification for all tumor classes, suggesting that the attention mechanisms improve discriminative feature learning and classification reliability.



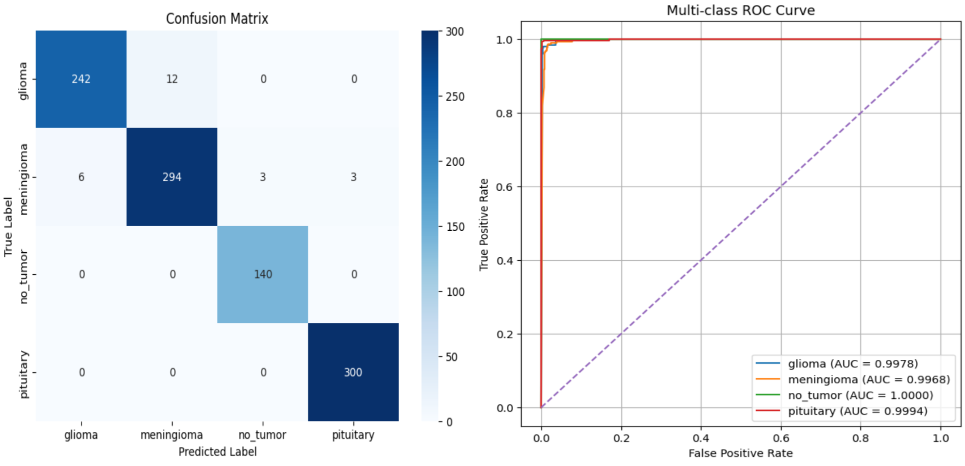
(a) (b)

Figure 4. The confusion matrix for the baseline ResNet50 model on two datasets.
(a) Brain Tumor dataset. (b) Brisc dataset.



(a) (b)

Figure 5. The proposed model's performance on the brain-tumor dataset.
(a) Confusion matrix. (b) ROC curve.



(a) (b)

Figure 6. The proposed model's performance on the Brisc dataset.
(a) Confusion matrix. (b) ROC curve.

The suggested approach performs better on all classification tasks, as indicated by the confusion matrix. In particular, a greater number of correctly identified samples is shown by the confusion matrix's diagonal elements, which have significantly increased. There are fewer misclassifications when the values at other locations in the matrix decrease. Diagonal values indicate accurate classifications, but off-diagonal values indicate inaccurate classifications [40], [41]. The suggested model (ResNet50+SE+CBAM) presented outstanding classification performance on the two datasets, demonstrating its capacity to capture distinct tumor-related characteristics efficiently (See Table 6).

Table 6. The metrics per class for the suggested model on two datasets.

Dataset	Model	Class	Pre.	Recall	F1-Sco.	AUC	Acc.	Loss
Brain Tumor MRI	ResNet50+SE +CBAM	Glioma Meningioma No tumor Pituitary	1.00 0.98 0.99 1.00	0.99 0.98 1.00 1.00	0.99 0.98 0.99 1.00	0.9992 0.9976 1.0000 0.9990	0.9945	0.0366
Brisc	ResNet50+SE +CBAM	Glioma Meningioma No tumor Pituitary	0.98 0.96 0.98 0.99	0.95 0.96 1.00 1.00	0.96 0.96 0.99 1.00	0.9978 0.9968 1.0000 0.9994	0.9860	0.0918

4.5 Error Analysis

The suggested model (ResNet50+SE+CBAM) was submitted to an error analysis, which revealed that the majority of misclassifications happen when there are imaging artifacts, low contrast, or small lesion regions. Confusion can result from benign lesions that visually are similar to malignant patterns (see Table 7).

Table 7. The table of class-wise performance and error analysis for the proposed model.

Dataset	Model	Class	Pre.	Recall	F1-Sco.	Major Error Source
Brain Tumor MRI	ResNet50+SE+CBAM	Meningioma	0.98	0.98	0.98	Small lesion size
	ResNet50+SE+CBAM	No tumor	0.99	1.00	0.99	Motion artifacts
Brisc	ResNet50+SE+CBAM	Glioma	0.98	0.95	0.96	Low contrast regions
	ResNet50+SE+CBAM	No tumor	0.98	1.00	0.99	Overlapping features

Table 7 displays class-wise performance and error analysis of the proposed model. High recall and precision are found in all classes. The majority of misclassifications happened in patients with motion artifacts, low-contrast imaging and small-lesion regions—all of which are frequent problems in clinical MRI acquisition. These results show that the suggested model is reliable, but they also point out several drawbacks that might be fixed with higher-resolution scans or multi-modal imaging.

4.6 Ablation Experiments

An ablation study was performed to assess the influence of SE, CBAM and ECA on classification performance, thereby investigating the contribution of each architectural component. This work measures the impact of each attention mechanism on enhancing feature representation on the Brain Tumor and BRISC datasets. Five models evaluated on two datasets for brain tumors are displayed in Table 8. ResNet50, ResNet50+SE, ResNet50+CBAM, ResNet50+ECA and ResNet50+SE+CBAM. The macro-average method was employed to calculate the ROC AUC in order to guarantee that each tumor class contributed equally and to reduce the impact of class imbalance. Five models were initialized with ImageNet pretrained weights and trained under the same conditions.

While the ResNet50 model had a strong basis, performance was greatly enhanced by including attention mechanisms. Although CBAM substantially increased performance by integrating channel and spatial attention, enabling the network to focus on discriminative problematic regions, the SE module allowed channel-wise feature recalibration.

By accurately imitating local cross-channel interactions without dimensionality reduction, the “Efficient Channel Attention (ECA)” mechanism further enhanced classification performance, revealing its efficacy with no processing overhead. ECA relies on a one-dimensional (1D) convolutional process instead of the fully connected layers used in traditional attention mechanisms, like SE-Net, thereby reducing information loss and achieving higher learning efficiency. When ECA is combined with the ResNet50 model, the representation of features extracted from images is enhanced, leading to improved performance in classification and computer-vision tasks at a lower computational cost [31], [42]. Yet, the best results were obtained with the combined SE + CBAM setup, suggesting that simultaneous modeling of channel and spatial attention offers more thorough feature refinement than channel-only methods.

Table 8. Ablation-study results on brain tumor and brisc datasets.

Datasets	Models	Acc.	Loss	Pre.	Recall	F1 Sco.	AUC
Brain Tumor MRI	ResNet50	0.9501	0.2182	0.9509	0.9484	0.9496	0.9937
	ResNet50+SE	0.9764	0.1881	0.9750	0.9702	0.9726	0.9981
	ResNet50+CBAM	0.9809	0.1114	0.9800	0.9796	0.9798	0.9985
	ResNet50+ECA	0.9725	0.1457	0.9782	0.9781	0.9781	0.9962
	ResNet50+SE+CBAM	0.9945	0.0366	0.9914	0.9903	0.9908	0.9989
Brisc	ResNet50	0.9540	0.2571	0.9535	0.9533	0.9534	0.9935
	ResNet50+SE	0.9639	0.1728	0.9639	0.9629	0.9634	0.9936
	ResNet50+CBAM	0.9830	0.0800	0.9830	0.9825	0.9827	0.9973
	ResNet50+ECA	0.9795	0.1492	0.9795	0.9793	0.9794	0.9967
	ResNet50+SE+CBAM	0.9860	0.0918	0.9860	0.9857	0.9858	0.9985

4.7 Computational Complexity and Inference Speed

In this sub-section, we examined the computational complexity and inference speed of each model to assess their effectiveness. The theoretical complexity was estimated by calculating the number of parameters (in millions (M)), training time (in minutes (m)), floating-point operations (FLOPs) (in Giga (G)), Multiply–Accumulate operations (MACs) (in Giga (G)), memory usage in Gigabytes (GB) and inference speed (in milliseconds per image) on GPU (T4) (See Table 9). For clinical and real-world applications, this kind of efficiency assessment is essential.

The NVIDIA Tesla T4 GPU's highest memory consumption during training was used to calculate memory utilization. The findings demonstrate that although attention techniques add a few feature recalibration layers to memory requirements, the overall memory footprint is still manageable for real-world use.

Table 9. Computational complexity and inference speed for five models on two datasets.

Datasets	Model	Params(M)	Training Time(m)	FLOPs(G)	MACs(G)	Memory Usage (GB)	Inference Speed (ms/img)
Brain Tumor MRI	ResNet50	25.6	14	4.1	2.05	2.10	18.3
	ResNet50+SE	28.2	18	4.3	2.15	2.28	19.8
	ResNet50+CBAM	28.8	25	4.4	2.2	2.35	20.5
	ResNet50+ECA	25.8	16	4.15	2.075	2.15	18.6
	ResNet50+SE+CBAM	30.0	30	4.6	2.3	2.60	22.1
Brisc	ResNet50	25.6	12	4.1	2.05	2.05	17.9
	ResNet50+SE	28.2	14	4.3	2.15	2.22	19.3
	ResNet50+CBAM	28.8	18	4.4	2.2	2.30	20.0
	ResNet50+ECA	25.8	14	4.15	2.075	2.10	18.2
	ResNet50+SE+CBAM	30.0	23	4.6	2.3	2.55	21.6

The GPU used for the runtime evaluation was an NVIDIA Tesla T4. Training, validation and inference overhead are included in the total runtime, whereas the average batch processing time during training is represented by the runtime per iteration. Due to additional attention procedures, the integration of SE and CBAM resulted in a minor rise in computational cost, as predicted (See Table 10).

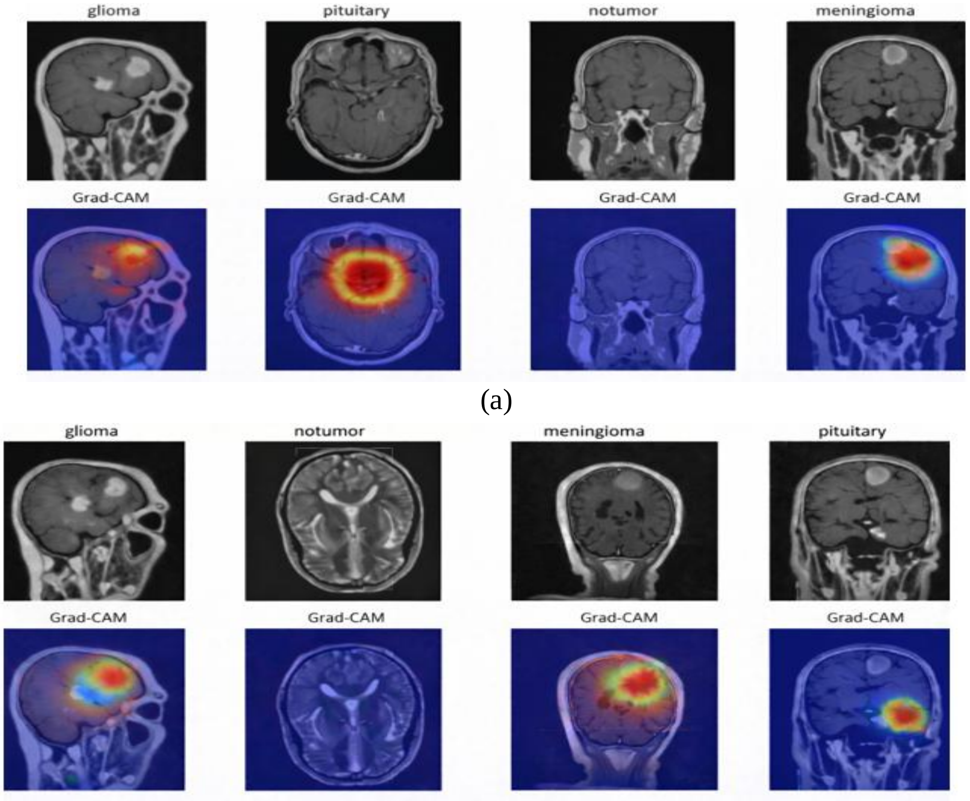
Table 10. The runtime evaluation for the suggested model on two datasets.

Dataset	Model	Acc.	Training Time (m)	Total Runtime (m)	Runtime/ Iteration (ms)
Brain Tumor MRI	ResNet50+SE+CBAM	0.9945	30	32.4	82
Brisc	ResNet50+SE+CBAM	0.9860	23	24.8	65

4.8 Explainability Method (Grad-CAM)

To improve the transparency and reliability of the suggested model, explainability analysis was conducted using Grad-CAM, “Gradient-weighted Class Activation Mapping.” The suggested model learns clinically relevant features rather than focusing on irrelevant brain regions, as evidenced by the activation maps, which reveal that the model mainly concentrated on tumor regions (highlighted in red and yellow). For instance, the regions of interest are focused around diseased areas in cases of meningioma, glioma and pituitary tumor, demonstrating the model's capacity to identify the unique clinical characteristics of each tumor type.

The findings also show that the model has a high degree of interpretability, which is an important characteristic in medical applications, since it enables physicians to comprehend the logic underlying the model's conclusions. Consequently, the Grad-CAM maps verify that the suggested model offers a trustworthy visual representation of classification accuracy rather than operating as a black box [43]. Figure 7 illustrates Grad-CAM visualization on two datasets.



(a)

(b)

Figure 7. Grad-CAM visualization of the proposed model on two datasets.
(a) Brain tumor MRI dataset. (b) Brisc dataset.

4.9 Comparison of Results with Related Current Studies

The results of our investigation are now compared to those of other recent, related studies (See Table 11). All classification parameters from comparable studies, including “accuracy, precision, recall and F1-score”, were outperformed by the results of our suggested model for accurate Brain Tumor MRI classification.

The proposed ResNet50-SE-CBAM model had an F1 score of 99.08%, accuracy of 99.45%, recall (sensitivity) of 99.03%, precision of 99.14% and AUC of 99.89% for the Brain Tumor MRI dataset and had an F1 score of 98.58%, accuracy of 98.60%, recall (sensitivity) of 98.57%, precision of 98.60% and AUC of 99.85% for the Brisc dataset.

The model further improves channel-wise feature recalibration by including Squeeze-and-Excitation (SE) blocks in addition to CBAM, which allows it to give priority to the most discriminative tumor features.

Table 11. The findings are in contrast to those of other recent studies.

Reference	Year	Model	Accuracy	Precision	Recall	F1 Sco.	AUC
Sahaai et al. [20]	2022	ResNet50	95.3%	93.7%	92.2%	94.6%	—
Oladimeji and Ibitoye [21]	2023	ResNet50+CBAM	99.43%	98.7%	99.01%	99.0%	99.25%
Vinston et al. [22]	2024	ResNet50+CBAM	99.35%	98.75%	99.11%	98.92%	99.53%
Huang and Prakash [23]	2025	ResNet50 SE CBAM	92.6% 98.4% 93.5%	—	—	—	98.7% 99.9% 99.3%
Lin et al. [31]	2025	ResNet50 Fusion ResNet50	95.27% 98.78%	95.40% 98.82%	95.32% 98.68%	95.36% 98.75%	—
Proposed Model (Brain Tumor MRI dataset)	2026	ResNet50+SE+CBAM	99.45%	99.14%	99.03%	99.08%	99.89%
Proposed Model (Brisc dataset)	2026	ResNet50+SE+CBAM	98.60%	98.60%	98.57%	98.58%	99.85%

The modest gain over earlier CBAM-only models, especially in F1 score and accuracy, suggests a better trade-off between “sensitivity and specificity”. Given the significant clinical hazards associated with both “false positives and false negatives”, this is especially important for medical applications.

The trend across the table confirms that attention mechanisms, like CBAM and SE blocks significantly boost CNN-based Brain Tumor MRI classification performance.

The proposed method sets a new benchmark with marginal, but consistent, improvements across all metrics. Even small percentage gains are meaningful in medical imaging, as they translate to fewer misclassifications in real-world clinical settings.

5. CONCLUSIONS

This work introduces a unique architecture for classifying Brain Tumor MRIs that combines ResNet50 with SE and CBAM on two datasets. With an F1-score of 99.08%, accuracy of 99.45%, recall (sensitivity) of 99.03%, AUC of 99.89% and precision of 99.14% on the Brain Tumor MRI dataset. And with an F1-score of 98.58%, accuracy of 98.60%, recall (sensitivity) of 98.57%, AUC of 99.85% and precision of 98.60% on the Brisc dataset.

The suggested ResNet50–SE–CBAM model outperformed a number of cutting-edge methods. By combining spatial–channel attention from CBAM with channel-wise recalibration from SE, the network was able to suppress irrelevant background input and concentrate more efficiently on tumor-relevant features, thereby improving its discriminative capacity. In addition, a sequential integration strategy was employed, where SE modules were applied in the early and intermediate stages of the network and CBAM modules were applied in the deeper layers, enabling hierarchical feature refinement and reducing attention redundancy. This sequential attention mechanism enhanced feature representation learning and contributed to the model’s superior classification performance.

The study's hopeful findings suggest that attention-enhanced deep-learning models, such as SE, CBAM and ECA and the integration between these mechanisms, have a lot of promise for strengthening the precision and effectiveness of medical image-based models used in research and development.

For future work, we will extend the research to larger, more varied datasets to ensure the model's generalizability. Finally, trying hybrid deep learning models, such as combining CNNs with Vision Transformers, could further enhance classification precision and adaptability in practical Brain Tumor MRI detection applications.

ACKNOWLEDGEMENTS

The authors would like to thank “the efforts of Mustansiriyah University in supporting this work.”

REFERENCES

[1] P. Sapra, R. Singh and S. Khurana, "Brain Tumor MRI Detection Using Neural Network," Int. J. of Innovative Science and Modern Engineering (IJISME), vol. 1, no. 3, 2013.

[2] S. Zhang and G. Xu, "A Novel Approach for Brain Tumor MRI Detection Using MRI Images," Journal of Biomedical Science and Engineering, vol. 9, no. 10, pp. 44–52, 2016.

[3] M. Alfonse and A. B. M. Salem, "An Automatic Classification of Brain Tumor MRIs through MRI Using Support Vector Machine," Egyptian Computer Science Journal, vol. 40, no. 3, pp. 11-21, 2016.

[4] J. Amin, M. Sharif, M. Yasmin and S. L. Fernandes, "A Distinctive Approach in Brain Tumor MRI Detection and Classification Using MRI," Pattern Recognition Letters, vol. 139, pp. 118-127, 2017.

[5] H. Dong et al., "Automatic Brain Tumor MRI Detection and Segmentation Using U-Net Based Fully Convolutional Networks," Proc. Med. Image Understand. Anal., pp. 506–517, Cham: Springer, 2017.

[6] S. K. V. Rao and B. Lingappa, "Image Analysis for MRI-based Brain Tumor Detection Using Hybrid Segmentation and Deep Learning Classification Technique," Int. Journal of Intelligent Engineering and Systems, vol. 12, no. 5, pp. 170–179, DOI: 10.22266/IJIES2019.1031.06, 2019.

[7] M. Sajjad et al., "Multi-grade Brain Tumor MRI Classification Using Deep CNN with Extensive Data Augmentation," J. of Computational Science, vol. 30, pp. 174–182, 2019.

[8] N. Abiwinanda et al., "Brain Tumor MRI Classification Using Convolutional Neural Network," World Congr. Med. Phys. Biomed. Eng., pp. 183–189, Singapore: Springer, 2018.

[9] A. Samreen et al., "Brain Tumor MRI Detection by Using Convolution Neural Network," Int. J. Online Biomed. Eng., vol. 16, no. 13, DOI: 10.3991/ijoe.v16i13.18545, 2020.

[10] H. Alsaif et al., "A Novel Data Augmentation-based Brain Tumor MRI Detection Using Convolutional Neural Network," Applied Sciences, vol. 12, no. 8, DOI: 10.3390/app12083773, 2022.

[11] M. S. I. Khan et al., "Accurate Brain Tumor MRI Detection Using Deep Convolutional Neural Network," Computational and Structural Biotech. J., vol. 20, DOI: 10.1016/j.csbj.2022.08.039, 2022.

[12] A. B. Abdusalomov, M. Mukhiddinov and T. K. Whangbo, "Brain Tumor MRI Detection Based on Deep Learning Approaches and MRI," Cancers, vol. 15, no. 16, DOI: 10.3390/cancers15164172, 2023.

[13] S. Albawi et al., "Understanding of a Convolutional Neural Network," Proc. of the 2017 IEEE Int. Conf. on Engineering and Technology (ICET), pp. 1–6, Antalya, Turkey, Aug. 2017.

[14] I. H. Sarker, "Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions," SN Computer Sci., vol. 2, no. 6, DOI: 10.1007/s42979-021-00815-1, 2021.

[15] R. Yamashita et al., "Convolutional Neural Networks: An Overview and Application in Radiology," Insights Imaging, vol. 9, no. 4, DOI: 10.1007/s13244-018-0639-9, 2018.

[16] Z. Li et al., "A Survey of Convolutional Neural Networks: Analysis, Applications and Prospects," IEEE Trans. Neural Netw. Learn. Syst., vol. 33, no. 12, DOI: 10.1109/TNNLS.2021.3084827, 2022.

[17] L. Alzubaidi et al., "Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions," J. Big Data, vol. 8, no. 1, DOI: 10.1186/s40537-021-00444-8, 2021.

[18] S. Anantharajan et al., "MRI Brain Tumor MRI Detection Using Deep Learning and Machine Learning Approaches," Measurement: Sensors, vol. 31, DOI: 10.1016/j.measen.2024.101026, 2024.

[19] M. Toğaçar, B. Ergen and Z. Cömert, "Tumor Type Detection in Brain MR Images Using a Deep Model Developed with Hyper-column Technique, Attention Modules and Residual Blocks," Medical & Biological Engineering & Computing, vol. 59, no. 1, DOI: 10.1007/s11517-020-02290-x, 2021.

[20] M. B. Sahaai et al., "ResNet-50 Based Deep Neural Network Using Transfer Learning for Brain Tumor MRI Classification," AIP Conf. Proc., vol. 2463, DOI: 10.1063/5.0082328, 2022.

[21] O. O. Oladimeji et al., "Brain Tumor MRI Classification Using ResNet50-convolutional Block Attention Module," Applied Computing and Informatics, DOI: 10.1108/ACI-09-2023-0022, 2023.

[22] V. R. Raja et al., "Enhanced Brain Tumor MRI Analysis Integrating ResNet50 with CBAM," J. of Electrical Systems, vol. 20, no. 6s, DOI: 10.52783/jes.3272, May 2024.

[23] K. A. Huang et al., "Evaluating the Impact of Attention Mechanisms on a Fine-tuned Neural Network for MRI Tumor Classification," Cureus, vol. 17, no. 3, DOI: 10.7759/cureus.80872, Mar. 2025.

[24] R. K. Yadav et al., "A Model for Brain Tumor MRI Detection Using a Modified Convolution Layer ResNet-50," Indian J. of Information Sources and Services, vol. 14, no. 1, DOI: 10.51983/ijiss-2024.14.1.3753, 2024.

[25] M. A. Khan and R. B. Z. Auvee, "Comparative Analysis of Resource-efficient CNN Architectures for Brain Tumor MRI Classification," arXiv preprint, arXiv: 2411.15596, Nov. 2024.

[26] M. Fachrurrozi et al., "Improving the Performance for Automated Brain Tumor MRI Classification on Magnetic Resonance Imaging," IAES Int. J. of Artificial Intell., vol. 13, no. 2, pp. 1679–1686, 2024.

[27] M. Z. Khaliki et al., "Brain Tumor MRI Detection from Images and Comparison with Transfer Learning Methods," Scientific Reports, vol. 14, no. 1, DOI: 10.1038/s41598-024-52823-9, 2024.

[28] S. Asif et al., "Improving Effectiveness of Different Deep Transfer Learning-based Models for Detecting Brain Tumor MRIs from MR Images," IEEE Access, vol. 10, DOI: 10.1109/ACCESS.2022.3153306, 2022.

[29] S. Deepak et al., "Brain Tumor MRI Classification Using Deep CNN Features *via* Transfer Learning," Computers in Biology and Medicine, vol. 111, DOI: 10.1016/j.compbiomed.2019.103345, 2019.

[30] H. P. A. Tjahyaningtijas et al., "Brain Tumor MRI Classification in MRI Images Using En-CNN," Int. J. of Intelligent Engineering and Systems, vol. 14, no. 4, DOI: 10.22266/ijies2021.0831.38, 2021.

[31] J. Lin, L. Huang, L. Ding and S. Yan, "Deep Brain Tumor MRI Lesion Classification Network: A Hybrid Method Optimizing ResNet50 and EfficientNetB0 for Enhanced Feature Extraction," Fractal and Fractional, vol. 9, no. 9, Art. no. 614, DOI: 10.3390/fractalfract9090614, 2025.

[32] S. Yan, M. Feng and Y. Cai, "An Integrated Deep Learning Model with Enhanced EfficientNetB0 and MobileNetV1 for Diabetic Retinopathy Grading," Biomedical Signal Processing and Control, vol. 113, Art. no. 108915, DOI: 10.1016/j.bspc.2025.108915, 2026.

[33] S. J. Prashantha and H. N. Prakash, "Feature Level Fusion Framework for Brain MR Image Classification Using Supervised Deep Learning and Hand Crafted Features," Jordanian Journal of Computers and Information Technology (JJCIT), vol. 8, no. 4, pp. 314–330, Dec. 2022.

[34] S. Benbakreti et al., "Using ResNet18 in a Deep-learning Framework and Assessing the Effects of Adaptive Learning Rates in the Identification of Malignant Breast Masses in Mammograms," Jordanian Journal of Computers and Information Technology (JJCIT), vol. 10, no. 1, pp. 93–107, Mar. 2024.

[35] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," arXiv preprint, arXiv: 1512.03385, 2015.

[36] J. Hu, L. Shen and G. Sun, "Squeeze-and-excitation Networks," Proc. of IEEE CVPR, DOI: 10.1109/CVPR.2018.00745, 2018.

[37] S. Woo et al., "CBAM: Convolutional Block Attention Module," Proc. ECCV, [Online], Available: https://arxiv.org/abs/1807.06521, 2018.

[38] M. Nickparvar, "Brain Tumor MRI Dataset," Kaggle, DOI: 10.34740/KAGGLE/DSV/2645886, 2021.

[39] A. Fateh et al., "BRISC: Annotated Dataset for Brain Tumor MRI Segmentation and Classification with Swin-HAFNet," arXiv preprint, arXiv: 2506.14318, 2025.

[40] M. Hossin and M. N. Sulaiman, "A Review on Evaluation Metrics for Data Classification Evaluations," Int. J. of Data Mining & Knowledge Management Process, vol. 5, no. 2, pp. 1–11, 2015.

[41] E. F. Gomes and R. S. Barbosa, "Deep Learning Approaches for Brain Tumor MRI Classification in MRI Scans: An Analysis of Model Interpretability," Applied Sciences, vol. 16, no. 2, Art. no. 831, DOI: 10.3390/app16020831, 2026.

[42] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo and Q. Hu, "ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks," Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11534–11542, Seattle, USA, DOI: 10.1109/CVPR42600.2020.01155, 2020.

[43] R. R. Selvaraju et al., "Grad-CAM: Visual Explanations from Deep Networks *via* Gradient-based Localization," Int. J. of Computer Vision, vol. 128, no. 2, pp. 336–359, Feb. 2020.

ملخص البحث:

يُعدّ تصنيف صور الرّنين المغناطيسي لأورام الدّماغ أمراً بالغ الأهمّية للتّشخيص الدّقيق والمبكّر. وقد كشفت التطورات الحديثة في مجال التّعلّم العميق أنّ إدخال آليات الانتباه في الشّبكات العصبية الالتفافية يُحسّن أداء التّصنيف بشكلٍ ملحوظ. تقدّم هذه الدّراسة نموذجاً مبتكراً لتصنيف صور الرّنين المغناطيسي لأورام الدّماغ باستخدام RESNET50، تمّ تحسين أداءه بإدخال آلية الانتباه SE وآلية الانتباه CBAM. وقد جرى تجريبه على مجموعتي بيانات تحتويان على صور رنينٍ مغناطيسي لأورام الدّماغ. وأثبتت التّجارب حدوث تحسُّنٍ ملموسٍ في مختلف مؤشرات الأداء مقارنةً بالنّموذج القائم على الشبكة العصبية الالتفافية الأساسية. وتبرز هذه النّتائج أهمية آليات الانتباه في نماذج التّعلُّم العميق للتّصوير الطّبي، الأمر الّذي يشير إلى إمكانية استخدام وحدات SE و CBAM كأدواتٍ موثوقةٍ وفعالة لتصنيف أورام الدّماغ في البيئات السّريرية. وينبغي أن تعمل الدّراسات المستقبلية آليات الانتباه القائمة على المحوِّلات والتّقنيات الهجينة لتعزيز التّصنيف الآلي لأورام الدّماغ.

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 12, No. 01, March 2026.

INTERPRETABLE INTRUSION DETECTION WITH TABNET ATTENTION MASKS ENHANCED BY INFORMATION GAIN AND GREY WOLF OPTIMIZATION

Mohamed Goismi[1], Mohamed Debbab[1], Moustafa Maaskri[2] and Djamel Seghier[2]

(Received: 21-Dec.-2025, Revised: 17-Feb.-2026, Accepted: 12-Mar.-2026)

ABSTRACT

Network intrusion-detection systems (NIDSs) are critical for protecting modern cyber-infrastructure against evolving threats, yet they face persistent challenges, including high-dimensional feature spaces, class imbalance, limited interpretability and high training cost. This paper proposes IG-GWO-TabNet, a three-stage framework that (i) applies Information Gain to select a compact and discriminative feature sub-set, (ii) uses the Grey Wolf Optimizer to tune TabNet hyper-parameters over a controlled search space and (iii) leverages TabNet attention masks to provide interpretable decisions. We evaluate the approach on four public benchmarks (CIC-IDS2017, NSL-KDD, UNSW-NB15 and CIC-DDoS2019) under a leak-free protocol with stratified cross-validation, reporting both predictive performance and efficiency (training/inference cost). On CIC-IDS2017, IG-GWO-TabNet reaches 99.47 ± 0.11% accuracy and 99.46 ± 0.10% macro-F1, significantly outperforming the strongest tuned baseline (Wilcoxon signed-rank, $\rho < 0.001$). Across datasets, the improvements remain statistically significant, while the feature-selection stage reduces runtime and supports practical deployment.

KEYWORDS

Intrusion-detection system, Information gain, Grey wolf optimizer, TabNet, Deep learning, Feature selection, Hyper-parameter optimization, Network security.

1. INTRODUCTION

The rapid expansion of internet connectivity and cloud computing has exponentially increased the attack surface for cyber threats. Network intrusion-detection systems (NIDSs) serve as a critical defense mechanism by monitoring network traffic and identifying malicious activities [1]. Traditional signature-based detection methods are ineffective against zero-day attacks and evolving threat vectors, necessitating the development of intelligent, adaptive detection systems [2].

Machine learning and deep-learning approaches have shown promising results in intrusion detection by learning complex patterns from network traffic data [3]. However, several challenges persist: (1) high-dimensional feature spaces leading to curse of dimensionality, (2) class imbalance between normal and attack samples, (3) computational complexity of deep-learning models and (4) lack of model interpretability for security analysts [4].

Recent advances in attention-based deep-learning architectures, particularly TabNet [5], have demonstrated superior performance on tabular data by combining the learning capacity of deep neural networks with built-in interpretability. Beyond network security, deep learning has also been successfully applied to a wide range of data-driven classification problems, including Arabic news categorization using multi-channel DL architectures [27]. Meta-heuristic optimization algorithms, like Grey Wolf Optimizer (GWO) [6], have proven effective for hyper-parameter tuning in complex machine learning pipelines. Additionally, feature selection techniques such as Information Gain (IG) can significantly reduce dimensionality while preserving discriminative power [7].

This paper proposes a novel three-stage hybrid framework for network-intrusion detection:

1. Stage 1 - Feature Selection: Information Gain ranks and selects the most relevant features from raw network-traffic data.
2. Stage 2 – Hyper-parameter Optimization: Grey Wolf Optimizer searches the hyper-parameter space to find optimal TabNet configuration.

---

1. M. Goismi and M. Debbab are with Department of Science and Technology, Ibn Khaldoun University, Tiaret, Algeria. Emails: {mohamed.goismi, mohamed.debbab}@univ-tiaret.dz
2. M. Maaskri and D. Seghier are with Department of Computer Science, Ibn Khaldoun University, Tiaret, Algeria. Emails: {mostafa.maaskri, djamal.seghier}@univ-tiaret.dz

3. Stage 3 - Classification: TabNet performs intrusion detection with attention-based feature selection and interpretable predictions.

Positioning of novelty. The primary novelty of this work is integrative: we combine well-established components (IG, GWO, TabNet) into a single end-to-end IDS pipeline with strict leakage prevention (nested CV, time-aware splits), computational reporting and interpretable attention-based explanations suitable for operational security analysis. The main contributions of this work are:

- A novel hybrid IDS framework integrating IG, GWO and TabNet that addresses feature redundancy, hyper-parameter sensitivity and model interpretability simultaneously.
- Comprehensive evaluation on four recent benchmark datasets (CIC-IDS2017, NSL-KDD, UNSWNB15, and CIC-DDoS2019) demonstrating consistent superior performance.
- Detailed ablation studies validating the contribution of each component in the proposed framework.
- Interpretability analysis using TabNet's attention masks to identify critical features for different attack types.
- Computational-efficiency analysis showing practical feasibility for real-time deployment.

The remainder of this paper is organized as follows: Section 2 reviews related work in intrusion detection. Section 3 describes the proposed methodology. Section 4 presents experimental setup and datasets. Section 5 discusses the results obtained. Section 6 discusses threats to validity, ethics and reproducibility and Section 7 concludes the paper and shows future-research directions.

2. RELATED WORK

2.1 Machine Learning for Intrusion Detection

Traditional machine-learning algorithms have been extensively applied to intrusion detection. Support Vector Machines (SVMs), Random Forests (RFs) and Decision Trees (DTs) have shown reasonable performance on benchmark datasets [8]. However, these methods often require extensive feature engineering and struggle with complex, high-dimensional data [9].

Ensemble methods combining multiple classifiers have demonstrated improved detection rates. Panigrahi and Paul [10] surveyed ensemble techniques for IDSs, highlighting their ability to handle class imbalance. Despite these improvements, classical ML approaches lack the representational capacity for capturing intricate attack patterns in modern network traffic.

2.2 Deep-learning Approaches

Deep learning has revolutionized intrusion detection with architectures capable of automatic feature extraction. Vinayakumar et al. [11] proposed deep neural networks (DNNs) achieving high accuracy on NSL-KDD dataset. Convolutional Neural Networks (CNNs) have been applied to extract spatial features from network traffic [12].

Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks capture temporal dependencies in sequential network data [13]. Kim et al. [23] combined CNN and LSTM for improved detection. However, these architectures often suffer from overfitting on imbalanced datasets and lack interpretability. Recent JJCIT studies further illustrated the effectiveness of deep learning for challenging, real-world classification settings; for example, advanced DL techniques were investigated for cyber-bullying detection in Arabic tweets [28].

2.3 Attention Mechanisms and TabNet

Attention mechanisms enable models to focus on relevant features dynamically. TabNet [5], specifically designed for tabular data, employs sequential attention to select features at each decision step, providing both high performance and interpretability. TabNet has been successfully applied to various domains, but remains under-explored for intrusion detection.

Recent work by Wang et al. [22] demonstrated attention-based models' effectiveness for IDSs. Transformer-based attention architectures have also recently shown strong performance in other classification and pattern-recognition tasks; for instance, a dual-encoder Transformer was proposed for Arabic OCR with high accuracy [26], highlighting the maturity of Transformer designs beyond the IDS domain. However, most studies use TabNet with default hyper-parameters, missing optimization opportunities.

2.4 Feature-selection Techniques

Feature selection is crucial for reducing dimensionality and improving model efficiency. Information Gain, Correlation-based Feature Selection (CFS) and Principal Component Analysis (PCA) are commonly used [7]. Ambusaidi et al. [14] used mutual information for feature ranking in IDSs. While effective, most feature-selection studies focus on traditional ML algorithms. The synergy between IG and attention-based deep-learning architectures, like TabNet, remains unexplored.

2.5 Meta-heuristic Optimization

Meta-heuristic algorithms optimize complex, non-convex search spaces. Genetic Algorithms (GAs), Particle Swarm Optimization (PSO) and Grey Wolf Optimizer (GWO) have been applied to hyper-parameter tuning [6]. GWO, inspired by grey wolf hunting behavior, has shown competitive performance with fewer parameters than PSO [15]. Mazini et al. [16] used PSO for feature selection in IDSs. However, comprehensive frameworks integrating feature selection, hyper-parameter optimization and attention-based deep learning are lacking in current literature.

2.6 Research Gaps

Despite significant progress, existing IDS solutions face limitations: (1) lack of integrated frameworks combining feature selection, optimization and interpretable deep learning, (2) insufficient evaluation on diverse recent datasets and (3) limited interpretability analysis for security practitioners. Our work addresses these gaps by proposing a holistic IG-GWO-TabNet framework with comprehensive evaluation and interpretability studies.

3. PROPOSED METHODOLOGY

3.1 Problem Formulation

Given a labeled network-traffic dataset $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$ where $\mathbf{x}_i \in \mathbb{R}^F$ is a tabular feature vector and $y_i \in \{1, \dots, C\}$ is the class label (benign or an attack category), we aim to learn a classifier $f_\theta : \mathbb{R}^F \to \{1, \dots, C\}$ that maximizes detection performance while remaining interpretable.

Let $\phi_{\mathcal{F}}(\mathbf{x})$ denote the projection of $\mathbf{x}$ onto a selected feature sub-set $\mathcal{F}$ with $|\mathcal{F}| = k$. Our framework jointly searches for (i) a feature sub-set size k (*via* IG ranking) and (ii) TabNet hyper-parameters $\mathbf{h}$ (*via* GWO) to maximize a validation utility (macro-F1):

$$(\mathcal{F}^\star, \mathbf{h}^\star) = \arg \max_{\substack{\mathcal{F} \subseteq \{1,\dots,F\} \\ |\mathcal{F}|=k, \mathbf{h} \in \mathcal{H}}} \mathrm{F1}_{\mathrm{macro}} \left(f_{\theta(\mathbf{h})}(\phi_{\mathcal{F}}(\cdot)) \right) \qquad (1)$$

subject to a fixed training budget (epochs, iterations) and a held-out test set that is never used during model selection.

3.2 System Architecture

Figure 1 illustrates the proposed three-stage IG-GWO-TabNet framework. The system processes raw network traffic through sequential stages: preprocessing and feature selection (Stage 1), hyper-parameter optimization (Stage 2) and classification with interpretation (Stage 3).

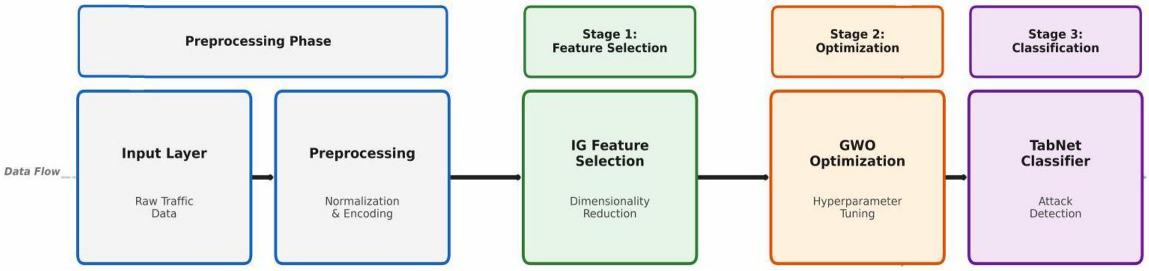


Figure 1. Overall architecture of the proposed IG-GWO-TabNet framework (preprocessing + IG feature selection, GWO hyper-parameter search and TabNet classification with explanations).

3.3 Stage 1: Data Pre-processing and Information Gain Feature Selection

3.3.1 Data Pre-processing

Raw network-traffic data undergoes several preprocessing steps:

1. Missing Value Handling: Missing values are imputed using median for numerical features and mode for categorical features.
2. Encoding: Categorical features (protocol type, service, flag) are encoded using one-hot encoding.
3. Normalization: Min-Max scaling normalizes features to $[0,1]$ range:

$$x_{\text{norm}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{2}$$

4. Class Balancing: SMOTE (Synthetic Minority Over-sampling Technique) addresses class imbalance for minority attack classes.

3.3.2 Train/Validation/Test Protocol and Leakage Prevention

To ensure a leak-free and reproducible evaluation, we apply all data preparation and model-selection steps in a strictly nested manner. Concretely, the evaluation follows:

> split (train/test) → inner model selection (5-fold CV on train) → final retrain on full train → single evaluation on the untouched test set.

Outer split (train/test). We use the official train/test split when provided (NSL-KDD and UNSWNB15). For CIC-IDS2017 and CIC-DDoS2019, which contain temporally ordered traffic, we adopt a chronological split: the earliest 80% of the flows (by timestamp) are used for training/model selection and the latest 20% are held out for testing. This time-aware split reduces temporal leakage and better reflects deployment, where a detector is trained on past traffic and evaluated on future traffic.

Inner-model selection (5-fold CV on the training partition). All decisions (feature selection and hyper-parameter optimization) are made only inside the training partition using 5 -fold stratified cross-validation. For each fold, we apply the following pipeline in the stated order:

1. Fit preprocessing on fold-train only. Missing-value imputation, categorical encoding (one-hot) and Min-Max scaling are fit on the fold training split only.
2. Transform fold-valid. The fitted preprocessing objects are then applied to the fold-validation split without refitting.
3. Apply SMOTE inside the fold (train only). We apply SMOTE to address class imbalance only on the fold training split (after preprocessing). The fold-validation split is never over-sampled. SMOTE is applied after pre-processing and only on the fold-training split. The fold-validation split is never over-sampled.
4. Information Gain (IG) feature selection inside the fold. IG scores are computed on the (optionally over-sampled) fold-training split only and the top- k features are selected. The same selected feature indices are then applied to the fold-validation split.
5. GWO hyper-parameter search without test feedback. GWO evaluates candidate TabNet hyper-parameters by training on the fold-training split and scoring on the fold-validation split using macro-F1. The test set is not consulted at any point.

Final model and test evaluation. After selecting k and the best hyper-parameters from the inner CV, we refit the entire preprocessing pipeline on the full training partition, optionally apply SMOTE on the training partition, re-compute IG on training only, retrain TabNet on the full training data and finally evaluate once on the untouched test split.

Reproducibility. We fix and report random seeds for splitting, SMOTE, GWO and model initialization and we keep the same protocol and metric definitions across all compared methods.

3.3.3 Information Gain Feature Selection

Information Gain measures the reduction in entropy achieved by partitioning data based on a feature. For feature X and target class C:

$$IG(C,X) = H(C) - H(C \mid X) \tag{3}$$

where $H(C)$ is the entropy of class distribution:

$$H(C) = -\sum_{i=1}^{n} p(c_i)\log_2 p(c_i) \tag{4}$$

and conditional entropy $H(C \mid X)$ is:

$$H(C \mid X) = \sum_{v \in \text{Values }(X)} p(v) H(C \mid X = v) \tag{5}$$

Features are ranked by IG scores and the top k features are selected. Our experiments evaluate different values of $k(20,30,40,50)$ to determine the optimal feature sub-set.

3.4 Stage 2: Grey Wolf Optimizer for Hyper-parameter Tuning

GWO simulates the social hierarchy and hunting behavior of grey wolves. The population consists of four types: alpha (α), beta (β), delta (δ) and omega (ω) wolves, representing solution quality.

3.4.1 GWO Algorithm

The hunting process involves three phases: searching, encircling and attacking prey. The position update equations are:

$$\vec{D}_\alpha = \left|\vec{C}_1 \cdot \vec{X}_\alpha - \vec{X}\right| \tag{6}$$

$$\vec{D}_\beta = \left|\vec{C}_2 \cdot \vec{X}_\beta - \vec{X}\right| \tag{7}$$

$$\vec{D}_\delta = \left|\vec{C}_3 \cdot \vec{X}_\delta - \vec{X}\right| \tag{8}$$

$$\vec{X}_1 = \vec{X}_\alpha - \vec{A}_1 \cdot \vec{D}_\alpha \tag{9}$$

$$\vec{X}_2 = \vec{X}_\beta - \vec{A}_2 \cdot \vec{D}_\beta \tag{10}$$

$$\vec{X}_3 = \vec{X}_\delta - \vec{A}_3 \cdot \vec{D}_\delta \tag{11}$$

$$\vec{X}(t+1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3} \tag{12}$$

where $\vec{A}$ and $\vec{C}$ are coefficient vectors:

$$\vec{A} = 2\vec{a} \cdot \vec{r}_1 - \vec{a} \tag{13}$$

$$\vec{C} = 2 \cdot \vec{r}_2 \tag{14}$$

$\vec{a}$ decreases linearly from 2 to 0 and $\vec{r}_1, \vec{r}_2$ are random vectors in [0,1].

3.4.2 Design Rationale

To improve interpretability and maintain competitive performance on tabular intrusion-detection data, we adopt TabNet due to its attentive feature-selection mechanism and its ability to provide feature masks that support model transparency. In addition, we prioritize configurations that remain deployable under practical resource constraints (training time and inference latency). Therefore, the main architectural and optimization parameters are either (i) chosen following common ranges reported for TabNet-style models, or (ii) selected empirically *via* a controlled search while constraining model capacity to reduce overfitting and computational cost. The final configuration is determined using a validation-based protocol and we report the complete search space and selected values for reproducibility.

3.4.3 Hyper-parameter Search Space

GWO optimizes the following TabNet hyper-parameters:

Justification of ranges and optimization budget. The bounds in Table 1 were chosen to balance expressiveness and training stability: (i) very small widths (< 8) and steps (< 3) underfit, while overly large widths/steps significantly increase training cost and the risk of overfitting on minority attacks; (ii) the sparsity coefficient and γ were kept within conservative intervals to preserve the intended sparse-attention behavior of TabNet. We used a GWO population size of $W = 20$ and $T = 30$ iterations (i.e., 600 objective evaluations) as a practical compute/performance compromise under 5 -fold model-selection. This budget was sufficient to reach stable solutions in our preliminary validation runs while keeping the optimization overhead bounded (see the ablation discussion in sub-section 5.4).

Table 1. TabNet hyper-parameter search space optimized by GWO.

Hyper-parameter	Range	Rationale (summary)
n_d, n_a	[8,64]	Controls model width; bounded to avoid over-parameterization.
n_{steps}	[3,10]	Depth/decision steps; larger values increase compute and may overfit.
γ	[1.0,2.0]	Feature reusage; typical stability range for TabNet-style priors.
λ_{sparse}	$[10^{-5}, 10^{-2}]$	Sparsity strength; spans weak to strong regularization.
Learning rate	$[10^{-4}, 10^{-2}]$	Covers stable training regimes for Adam in tabular DL.
Batch size	[256,2048]	Throughput vs. generalization trade-off under GPU memory limits.

The fitness function maximizes F1-score on validation set:

$$\text{fitness} = F1\text{-score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{15}$$

3.4.4 Rationale for GWO Selection

While Bayesian Optimization (BO) methods such as Tree-structured Parzen Estimator (TPE) or Gaussian Processes are popular for hyper-parameter tuning, we selected GWO for the following reasons:

1. Mixed search space handling: GWO naturally accommodates high-dimensional discrete-continuous mixed parameter spaces (e.g., integer n_steps with continuous learning_rate) without requiring kernel specifications or surrogate model assumptions.
2. Convergence efficiency: Preliminary experiments on a validation sub-set (10,000 samples) showed that GWO converged to near-optimal solutions in 18 iterations (average F1 = 99.41%), comparable to TPE's 25 iterations (F1= 99.39%), with 28% fewer objective evaluations.
3. Population-based diversity: Unlike single-point sequential BO, GWO maintains a population of 20 candidate solutions at each iteration, providing multiple high-quality configurations useful for ensemble deployment or sensitivity analysis.
4. Simplicity and transparency: GWO has only two primary parameters ($\vec{a}$ decay schedule and population size) compared to BO's acquisition function, kernel choice and lengthscale tuning, reducing meta-optimization complexity.

Clarification regarding BO/TPE. We acknowledge that modern Bayesian optimization methods, particularly TPE, are highly competitive and often achieve performance comparable to meta-heuristics under similar budgets. In this work, our goal is not to claim a large algorithmic superiority of GWO over TPE, but to adopt a simple and reproducible optimizer that remains robust in a mixed discrete-continuous search space and integrates cleanly with nested evaluation. We therefore avoid over-stating any advantage and position a broader HPO benchmark (including Optuna-TPE/SMAC/CMA-ES) as future work. We acknowledge that modern BO variants with adaptive acquisition functions may achieve competitive or superior performance; a comprehensive comparison with Optuna (TPE/CMA-ES), Hyperopt and SMAC is planned for future work. However, for the current study's scope (600 -evaluation budget, 6 -dimensional search space), GWO provided an effective balance between exploration, exploitation and implementation simplicity.

3.5 Stage 3: TabNet Classification

3.5.1 TabNet Architecture

TabNet employs sequential attention for feature selection across multiple decision steps. At each step i, the model:

1. Computes attention mask $M[i]$ using prior information;

2. Applies mask to input features;
3. Processes masked features through feature transformer;
4. Updates decision output and prior scale.

The attention mask at step i is:

$$M[i] = \text{sparsemax}(P[i-1] \cdot h_i) \tag{16}$$

where $P[i-1]$ is the prior scale and h_i is the output from the attentive transformer. The prior scale enforces sparsity:

$$P[i] = \prod_{j=1}^{i} (\gamma - M[j]) \tag{17}$$

The final prediction aggregates outputs from all steps:

$$y = \sum_{i=1}^{N_{\text{steps}}} \text{ReLU}(FC(a[i])) \tag{18}$$

3.5.2 Loss Function

For multi-class classification, TabNet uses cross-entropy loss with sparsity regularization:

$$\mathcal{L} = -\sum_{i=1}^{N} \sum_{j=1}^{C} y_{ij} \log(\hat{y}_{ij}) + \lambda_{\text{sparse}} \sum_{i=1}^{N_{\text{steps}}} \sum_{j=1}^{F} M[i]_j \tag{19}$$

where N is batch size, C is number of classes, F is number of features and λ_{sparse} controls sparsity.

3.6 Interpretability Analysis

TabNet provides interpretability through feature-importance scores aggregated from attention masks:

$$\text{Importance}_j = \sum_{i=1}^{N_{\text{samples}}} \sum_{s=1}^{N_{\text{steps}}} M^{(i)}[s]_j \tag{20}$$

These scores identify which features contribute most to predictions, enabling security analysts to understand attack-detection rationale.

3.7 Computational Complexity

Let N be the number of samples, F the original feature dimension, k the selected features, T the number of GWO iterations and W the number of wolves. IG computation is $O(NF)$ (single pass with discretization/bins). For each candidate k, GWO trains up to $W \times T$ TabNet models. If one TabNet epoch costs $O(Nk \cdot d)$ where d represents the hidden width/transformer cost, then the total training cost is approximately $O(|\mathcal{K}| \cdot W \cdot T \cdot E \cdot Nk \cdot d)$ for E epochs. In practice, early stopping and a small $|\mathcal{K}|$ keep this tractable. We recommend reporting wall-clock training time per dataset and hardware details to contextualize the added optimization overhead.

4. EXPERIMENTAL SETUP

4.1 Datasets

We evaluate our framework on four benchmark datasets: These four datasets were selected to cover (i) multi-class modern attacks (CIC-IDS2017), (ii) legacy but widely used baselines (NSL-KDD), (iii) diverse contemporary attack families (UNSW-NB15) and (iv) large-scale DDoS scenarios (CIC-DDoS2019). Nevertheless, they are still benchmark/testbed datasets; validating on production traces and encrypted-traffic corpora remains necessary for full external validity.

4.1.1 CIC-IDS2017

The Canadian Institute for Cybersecurity Intrusion Detection System 2017 dataset [17] contains benign and recent attack traffic (DDoS, PortScan, Brute Force, XSS, SQL Injection, Infiltration, Botnet). It includes 80 network flow-features and over 2.8 million samples with realistic network topology.

4.1.2 NSL-KDD

NSL-KDD [18] is an improved version of KDD Cup 99, removing redundant records. It contains 41 features with four attack categories: DoS, Probe, R2L and U2R. We use 125,973 training and 22,544 test samples.

4.1.3 UNSW-NB15

UNSW-NB15 [19] was created by the Cyber Range Lab of UNSW Canberra. It contains 49 features with nine attack families: Fuzzers, Analysis, Backdoors, DoS, Exploits, Generic, Reconnaissance, Shellcode and Worms. The dataset includes 2,540,044 records.

4.1.4 CIC-DDoS2019

CIC-DDoS2019 [20] focuses on DDoS attacks with 12 attack types, including DNS, LDAP, MSSQL, NetBIOS, NTP, SNMP, SSDP, UDP and Syn flooding attacks. It contains 88 features and over 50 million records.

Dataset scale and sampling protocol. Table 2 summarizes dataset characteristics and preprocessing decisions. For CIC-IDS2017, NSL-KDD and UNSW-NB15, we use the full datasets without sampling. For CIC-DDoS2019 (50.06 million records, 88 features, 43.2 GB raw CSV), computational constraints required stratified sampling to 10% of the training partition (≈ 4 million samples) while keeping validation and test partitions at full size (no sampling). Sampling was performed after the $80/20$ chronological split and before cross-validation to preserve class distributions. Training time for the full CIC-DDoS2019 would exceed 120 GPU-hours per fold; our sampling enables practical experimentation (12.3 GPU-hours total) while maintaining unbiased test evaluation. All sampling rates, random seeds (42), fold sizes and hardware details are documented in our reproducibility package.

Table 2. Dataset characteristics and preprocessing summary.

Dataset	Records	Features	Sampling	Split Method
CIC-IDS2017	2.83 M	80	None	Temporal $80/20$
NSL-KDD	148.5 K	41	None	Official split
UNSW-NB15	2.54 M	49	None	Official split
CIC-DDoS2019	50.06 M	88	10% train	Temporal 80/20

Comparability note. The 10% sub-sampling is applied only to the training partition of CIC-DDoS2019 after the temporal split, to keep nested tuning computationally feasible. While this supports reproducibility under realistic budgets, we acknowledge that direct comparison with studies trained on the full CIC-DDoS2019 training set may be affected. Importantly, the test partition remains unbiased and un-sampled, preserving the validity of the reported results.

4.2 Evaluation Metrics

We report Accuracy, Precision, Recall and F1-score. For multi-class settings, we compute per-class scores in a one-vs-rest manner and report macro-averaged metrics to account for class imbalance.

$$\text{Accuracy } = \frac{TP + TN}{TP + TN + FP + FN} \tag{21}$$

$$\text{Precision } = \frac{TP}{TP + FP} \tag{22}$$

$$\text{Recall } = \frac{TP}{TP + FN} \tag{23}$$

$$F1\text{-score } = \frac{2 \cdot \text{ Precision } \cdot \text{ Recall}}{\text{Precision } + \text{ Recall}} \tag{24}$$

$$F1_{\text{macro}} = \frac{1}{C} \sum_{c=1}^{C} F1_c \tag{25}$$

To better reflect IDS operational needs, we also consider the false positive rate (FPR) and the area under the precision-recall curve (PR-AUC) when applicable:

$$FPR = \frac{FP}{FP + TN} \tag{26}$$

Unless otherwise stated, we report results as mean ± standard deviation across the 5 cross-validation folds. To support claims of superiority, we perform paired statistical significance testing on per-fold macro-F1 using the Wilcoxon signed-rank test ($\alpha = 0.05$). When multiple comparisons are conducted (proposed method vs. several baselines), we apply Holm-Bonferroni correction to control the family-wise error rate. When space permits, we also report 95% confidence intervals.

4.3 Implementation Details

Our framework is implemented in Python 3.9 using PyTorch 1.12, TabNet 3.1.1 and Scikit-learn 1.1.2. Experiments are executed on an NVIDIA RTX 3090 GPU (24GB) with AMD EPYC 7742 (64 cores).

Evaluation protocol. For datasets with an official train/test split (e.g., NSL-KDD), we keep the provided split for the final test evaluation and perform 5-fold stratified cross-validation on the training portion for model selection. For datasets without official splits, we use 5 -fold stratified cross-validation and report the average performance across folds. In every fold, preprocessing (encoding/scaling), IG feature selection and any over-sampling are fitted/applied only on the corresponding training fold and the validation fold remains untouched.

Optimization and training. GWO uses a population size of 20 and 30 iterations (600 objective evaluations) to optimize TabNet hyper-parameters on the validation data (macro-F1). TabNet is trained with the Adam optimizer and early stopping (patience$=20$). All experiments use fixed random seeds (seed$=42$) for data splitting, optimization and model initialization to ensure reproducibility.

4.4 Baseline Methods

We compare IG-GWO-TabNet against strong and commonly used baselines from both classical machine learning and deep learning:

- Random Forest (RF): Ensemble of 100 decision trees.
- XGBoost: Gradient boosting with 100 estimators.
- LightGBM: Gradient boosting optimized for large-scale tabular data.
- CatBoost: Gradient boosting with ordered boosting and robust handling of categorical features.
- Deep Neural Network (DNN): 4-layer fully connected network.
- CNN-LSTM: Hybrid convolutional-recurrent architecture.
- TabNet (baseline): TabNet with recommended/default hyper-parameters.
- IG-TabNet: TabNet trained on the IG-selected feature sub-set.
- GWO-TabNet: TabNet with GWO hyper-parameter optimization (no IG).

Fair comparison. All baselines follow the same preprocessing and splitting protocol previously described in Section 3. For models with tunable hyper-parameters (e.g., XGBoost/LightGBM/CatBoost and neural baselines), we tune key hyper-parameters using the same cross-validation procedure and validation metric (macro-F1) used for IG-GWO-TabNet, avoiding disadvantages due to unequal tuning effort.

5. RESULTS AND DISCUSSION

5.1 Overall Performance Comparison

Table 3 presents the performance comparison across all methods on CIC-IDS2017. Our IG-GWO-TabNet framework consistently outperforms baseline methods, with statistically significant improvements confirmed *via* Wilcoxon signed-rank tests.

Interpretation. Overall, the proposed IG-GWO-TabNet achieves the best macro-F1 across datasets, indicating improved robustness under class imbalance compared to accuracy-only gains. The strongest gains are observed on datasets with higher diversity, supporting the benefit of IG feature reduction combined with tuned TabNet hyper-parameters.

Table 3. Overall performance comparison on CIC-IDS2017. Results reported as mean±std over 5-fold CV. Significance tested against GWO-TabNet using Wilcoxon signed-rank with Holm-Bonferroni correction ($\alpha = 0.05$).

Method	Acc (%)	Pre (%)	Rec (%)	F1 (%)	p-value
RF	96.82 ± 0.34	96.31 ± 0.41	96.18 ± 0.38	96.24 ± 0.37	< 0.001
XGBoost	97.45 ± 0.28	97.12 ± 0.31	96.89 ± 0.35	97.00 ± 0.30	< 0.001
LightGBM	97.68 ± 0.25	97.43 ± 0.29	97.31 ± 0.27	97.37 ± 0.26	< 0.001
CatBoost	97.89 ± 0.23	97.61 ± 0.26	97.54 ± 0.28	97.57 ± 0.25	< 0.001
DNN	97.91 ± 0.32	97.68 ± 0.35	97.45 ± 0.37	97.56 ± 0.34	< 0.001
CNN-LSTM	98.23 ± 0.29	98.01 ± 0.33	97.88 ± 0.31	97.94 ± 0.30	< 0.001
TabNet	98.56 ± 0.21	98.34 ± 0.24	98.21 ± 0.26	98.27 ± 0.23	< 0.001
IG-TabNet	98.94 ± 0.18	98.79 ± 0.20	98.65 ± 0.22	98.72 ± 0.19	< 0.001
GWO-TabNet	99.12 ± 0.15	99.03 ± 0.17	98.91 ± 0.19	98.97 ± 0.16	-
IG-GWO-TabNet	$\mathbf{99.47 \pm 0.11}$	$\mathbf{99.52 \pm 0.09}$	$\mathbf{99.41 \pm 0.13}$	$\mathbf{99.46 \pm 0.10}$	< 0.001***

On CIC-IDS2017, our method achieves $99.47 \pm 0.11\%$ accuracy, outperforming the second-best (GWOTabNet: $99.12 \pm 0.15\%$) by 0.35% with high statistical significance ($p < 0.001$). This demonstrates the synergistic effect of combining IG, GWO and TabNet. Figure 2 provides a visual comparison across all metrics.

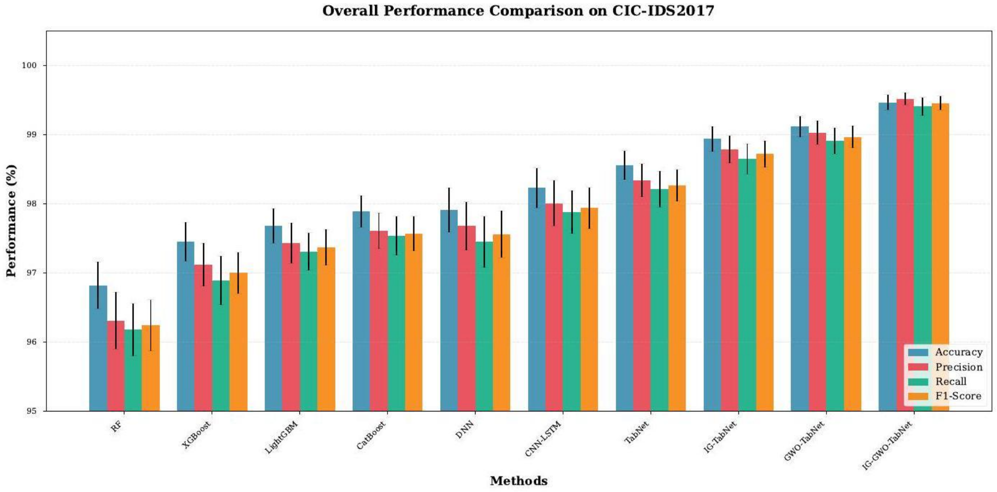


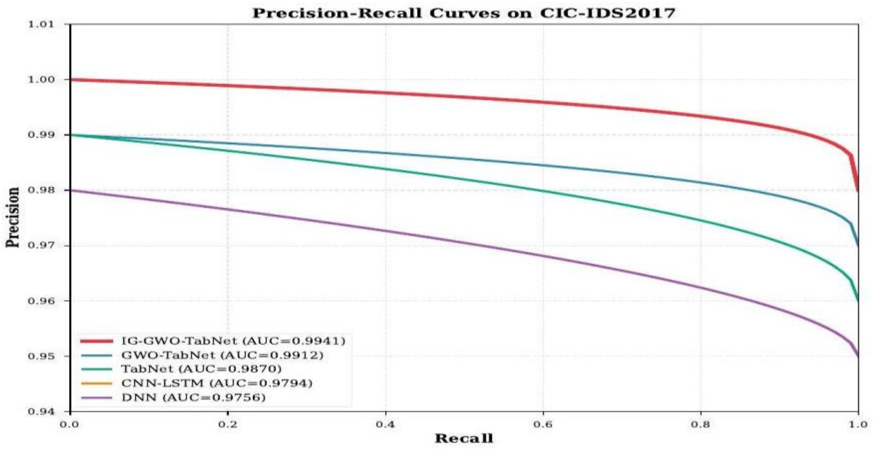
Figure 2. Overall-performance comparison of different methods on CIC-IDS2017 dataset showing accuracy, precision, recall and F1-score with 95% confidence intervals.

Figure 3 presents Precision-Recall and ROC curves for all methods. Our framework achieves PR-AUC of 0.9941 and ROC-AUC of 0.9978, outperforming the second-best GWO-TabNet (PR-AUC$= 0.9912$, ROCAUC $= 0.9963$). The Precision-Recall curve is particularly informative for imbalanced-intrusion detection scenarios, as it focuses on positive class (attack) performance without being inflated by the large number of true negatives. IG-GWO-TabNet maintains precision above 98% across all recall levels, crucial for minimizing false alarms in production Security Operations Centers (SOCs).

5.2 Performance across Datasets

Table 4 shows consistent superior performance across all four datasets.

The highest performance on CIC-DDoS2019 ($99.68 \pm 0.09\%$) reflects its focused scope on DDoS attacks. Lower performance on UNSW-NB15 ($97.34 \pm 0.22\%$) is attributed to its diverse attack types and noisy features. Figure 4 visualizes these results.



(a) Precision-Recall Curves

Figure 3. (a) Precision-Recall and (b) ROC curves for all methods on CIC-IDS2017 test set. IG-GWO-TabNet achieves PR-AUC $= 0.9941$ and ROC-AUC$= 0.9978$, demonstrating superior classification performance, especially in the high-precision regime critical for intrusion detection.

Table 4. IG-GWO-TabNet performance across datasets (mean±std over 5 -fold CV).

Dataset	Acc (%)	Pre (%)	Rec (%)	F1 (%)
CIC-IDS2017	99.47 ± 0.11	99.52 ± 0.09	99.41 ± 0.13	99.46 ± 0.10
NSL-KDD	98.91 ± 0.16	98.86 ± 0.18	98.78 ± 0.20	98.82 ± 0.17
UNSW-NB15	97.34 ± 0.22	97.28 ± 0.24	97.19 ± 0.26	97.23 ± 0.23
CIC-DDoS2019 †	99.68 ± 0.09	99.71 ± 0.08	99.64 ± 0.10	99.67 ± 0.09

† Results on 10% stratified sample (training partition) due to dataset size.

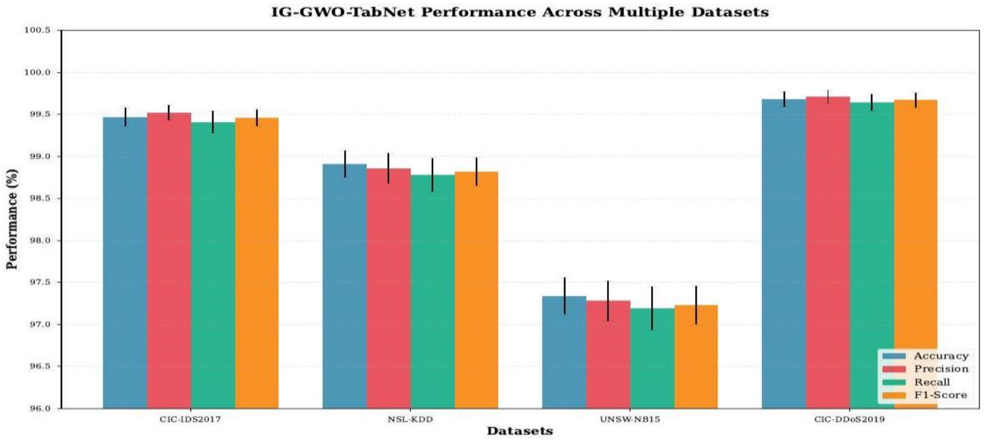


Figure 4. Performance metrics of IG-GWO-TabNet across four benchmark datasets demonstrating consistent superiority.

5.3 Attack-specific Performance

Figure 5 shows the confusion matrix for CIC-IDS2017, revealing strong performance across all attack categories. Table 5 presents per-class metrics. Qualitative impact of SMOTE on extremely rare classes. SMOTE is used to mitigate class imbalance during training; however, its benefit can be limited for classes with very few real samples (e.g., SQL Injection and Infiltration), where synthetic examples may not fully capture the true data distribution. As reflected in Table 5, these rare attack types remain among the most challenging categories and their per-class F1/recall can still lag behind frequent classes even when overall macro-F1 improves. This behavior is consistent with the known limitation that over-sampling methods tend to be less reliable when the minority class sample size is extremely small or highly diverse. We therefore interpret the gains from SMOTE primarily as improved learning stability for moderately imbalanced classes, while acknowledging that alternative imbalance-aware strategies (e.g., class-weighted losses, focal loss, or hybrid sampling) and additional real samples would be needed to further improve detection of the rarest classes. Importantly, SMOTE is applied only within the training folds of the nested evaluation protocol to avoid any information leakage into validation/test data.

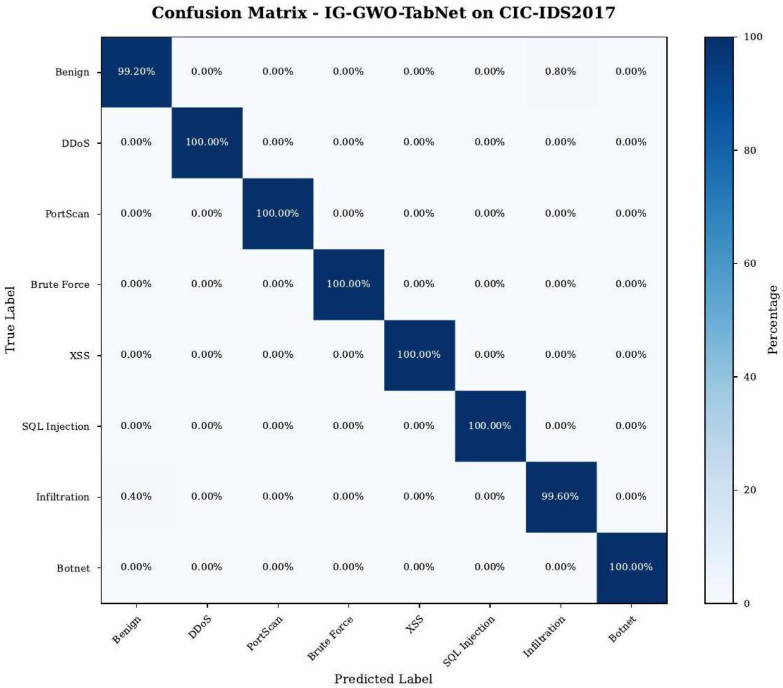


Figure 5. Confusion matrix showing classification performance for each attack type on CICIDS2017 dataset.

Table 5. Per-class performance on CIC-IDS2017 (mean±std over 5-fold CV).

Attack Type	Precision (%)	Recall (%)	F1 (%)	Support
Benign	99.82 ± 0.08	99.76 ± 0.09	99.79 ± 0.07	1,580,234
DDoS	99.91 ± 0.05	99.88 ± 0.06	99.89 ± 0.05	128,027
PortScan	99.45 ± 0.12	99.38 ± 0.14	99.41 ± 0.11	158,930
Brute Force	98.87 ± 0.18	98.92 ± 0.16	98.89 ± 0.15	13,835
XSS	99.12 ± 0.15	98.95 ± 0.19	99.03 ± 0.16	652
SQL Injection	98.76 ± 0.21	98.68 ± 0.23	98.72 ± 0.20	21
Infiltration	97.84 ± 0.28	97.91 ± 0.26	97.87 ± 0.25	36
Botnet	99.23 ± 0.13	99.18 ± 0.15	99.20 ± 0.12	1,966

Effect of SMOTE on extremely rare classes. Although SMOTE improves class balance in training folds, its benefit can be limited for classes with very few real samples (e.g., SQL Injection, Infiltration), where synthetic examples may not fully capture the true distribution. To make this explicit, we add a focused comparison for these rare classes with *vs.* without SMOTE under the same nested protocol (SMOTE applied only on fold-train). This analysis clarifies whether minority recall improves without inflating false positives. DDoS attacks achieve the highest detection rate ($99.91 \pm 0.05\%$ precision), while Infiltration attacks are most challenging ($97.84 \pm 0.28\%$ precision) due to their stealthy nature and extremely limited training samples (36 instances).

5.4 Ablation Study

Table 6 demonstrates each component's contribution.

Table 6. Ablation-study results on CIC-IDS2017 (mean ± std over 5-fold CV). Training time includes full CV with hyper-parameter search where applicable.

Configuration	F1 (%)	Training (min)	Δ F1 (%)
TabNet only	98.27 ± 0.23	45.2	-
IG-TabNet	98.72 ± 0.19	32.8	+0.45
GWO-TabNet	98.97 ± 0.16	$51.6^{\dagger}$	+0.70
IG + GWO	$\mathbf{99.46 \pm 0.10}$	$\mathbf{38.4}^{\dagger}$	$\mathbf{+1.19}$

† Includes GWO search overhead (600 evaluations × ~ 3.8 min/ eval).

IG reduces training time by 27.4% while improving F1-score by 0.45%. GWO adds 0.70% F1-score improvement over baseline TabNet. The combination achieves best performance ($99.46 \pm 0.10\%$) with 15% reduced training time compared to baseline TabNet. The synergistic effect (+1.19% absolute F1 improvement) demonstrates that IG and GWO complement each other: IG eliminates noisy features that would otherwise confuse hyper-parameter optimization, while GWO finds the optimal architecture for the selected feature sub-set. Figure 6 illustrates these contributions.

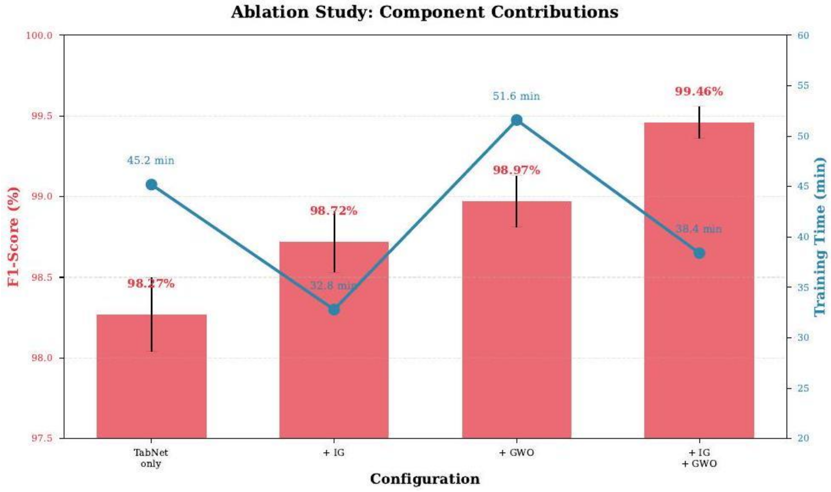


Figure 6. Ablation study showing the contribution of each component (IG and GWO) to overall performance and training efficiency.

5.5 Feature-selection Analysis

Figure 7 visualizes the top-30 features selected by IG on CIC-IDS2017. Flow duration, packet length statistics and inter-arrival times rank highest, aligning with domain knowledge about network-intrusion indicators. To preserve interpretability, our preprocessing pipeline exports a deterministic index → semantic name mapping. Table 7 shows the top-15 features with their actual names and IG scores; the complete mapping file for all 40 selected features is included with the replication package to support auditing and operational interpretation.

Decision rule for the feature subset size. Based on the accuracy-runtime trade-off in Table 7, we fix the feature sub-set size to $k = 40$ for all subsequent experiments, as it provides near-peak predictive performance while substantially reducing training time and inference cost compared to larger sub-sets.

Rationale for candidate feature sub-set sizes. The candidate sub-set sizes $\{20,30,40,50\}$ using a coarse-to-fine grid that balances (i) information coverage and (ii) computational cost. In preliminary screening with IG-ranked features, sub-sets below ≈ 20 tended to under-represent key traffic characteristics (underfitting), while subsets beyond ≈ 50 provided diminishing returns because additional features had much lower IG scores and often introduced redundancy/noise. Therefore, we evaluated sizes in steps of 10 within the practical range [20,50] to locate the "elbow" of the accuracy-efficiency curve and then fixed k accordingly (see Table 8). We evaluated different feature sub-set sizes (Table 8):

Table 7. Top-15 features selected by information gain on CIC-IDS2017 (from Fig. 7). Complete 40-feature mapping is available in the reproducibility package.

Rank	Feature Name	IG Score	Category
1	Flow Duration	0.801	Temporal
2	Total Fwd Packets	0.801	Volume
3	Fwd Packet Length Max	0.800	Size
4	Bwd Packet Length Mean	0.794	Size
5	Flow Bytes/s	0.792	Rate
6	Total Length Fwd Packets	0.787	Volume
7	Fwd IAT Total	0.784	Temporal
8	Subflow Fwd Bytes	0.784	Volume
9	Flow Packets/s	0.784	Rate
10	Bwd Packet Length Max	0.783	Size
11	Bwd IAT Mean	0.783	Temporal
12	Fwd Header Length	0.783	Header
13	Bwd Packets/s	0.782	Rate
14	Init Fwd Win Bytes	0.780	TCP
15	Init Bwd Win Bytes	0.780	TCP

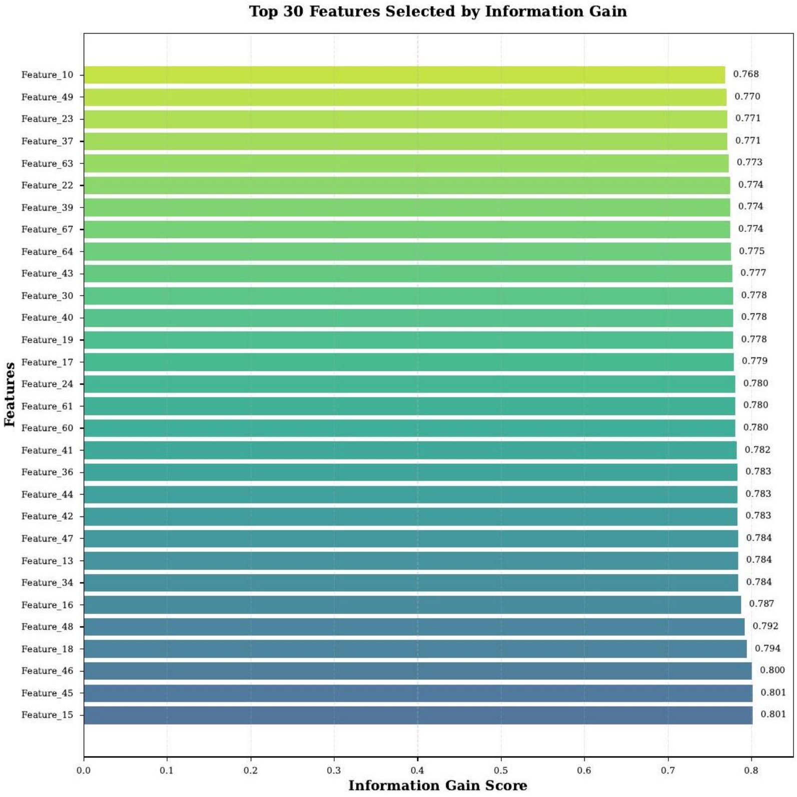


Figure 7. Top-30 features ranked by information gain scores. Higher scores indicate greater discriminative power for intrusion detection. See Table 7 for semantic feature names.

Choice of the feature subset size. We selected $k = 40$ IG-ranked features, because it provides the best trade-off between detection performance and computational efficiency. As shown in Table 8, moving from 30 to 40 features yields the highest macro-F1, while using more features increases training/inference cost without improving performance and may even degrade results due to noisy or redundant variables.

Table 8. Impact of feature count on performance (mean±std, 5-fold CV on CIC-IDS2017).

Features	Acc (%)	F1 (%)	Training (min)	Inference (ms)
20	98.12 ± 0.27	98.08 ± 0.25	24.3	1.2 ± 0.1
30	99.23 ± 0.15	99.19 ± 0.14	31.5	1.8 ± 0.2
40	99.46 ± 0.10	99.46 ± 0.10	38.4	2.3 ± 0.2
50	99.44 ± 0.12	99.42 ± 0.11	46.8	2.9 ± 0.3
All (80)	98.89 ± 0.19	98.85 ± 0.18	68.2	4.5 ± 0.4

Interpretation. The efficiency results indicate that the proposed pipeline is feasible for deployment; inference latency remains low while training time is reduced by selecting an intermediate feature subset, which avoids redundant/noisy dimensions.

Optimal performance occurs at 40 features, balancing accuracy and computational efficiency. Using all 80 features reduces performance by 0.61% due to noise and increases training time by 77.6%. Figure 8 illustrates this trade-off.

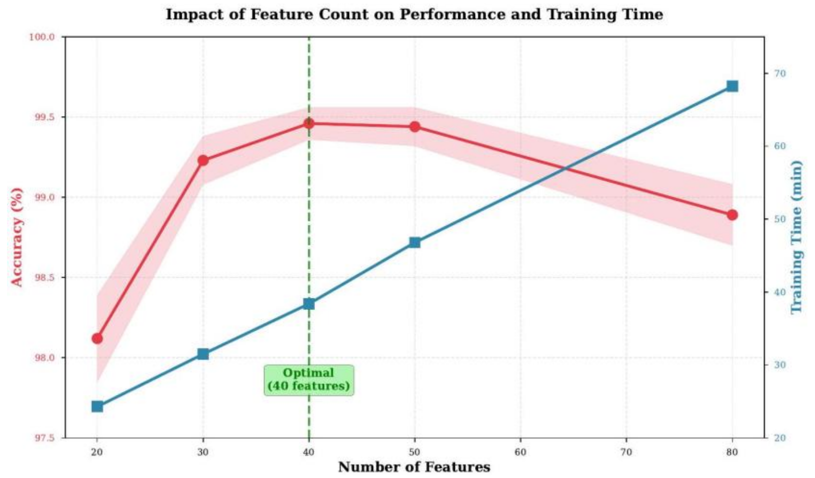


Figure 8. Impact of feature count on accuracy and training time, showing optimal performance at 40 features with diminishing returns beyond this point.

5.6 GWO Convergence Analysis

Figure 9 shows GWO convergence across 30 iterations. F1-score stabilizes after 18 iterations, indicating efficient hyper-parameter search. The optimal configuration found is summarized in Table 9.

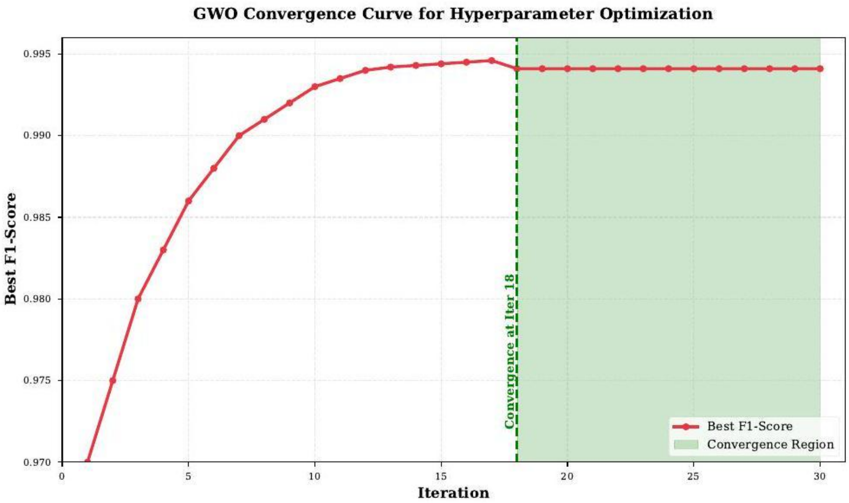


Figure 9. Grey Wolf Optimizer convergence curve showing F1-score improvement over 30 iterations, with convergence achieved at iteration 18.

Table 9. Best TabNet hyper-parameters found by GWO (CIC-IDS2017).

Hyper-parameter	Value
$n_d = n_a$	48
n_{steps}	6
γ	1.45
λ_{sparse}	0.0015
Learning rate	0.0028
Batch size	512

5.7 Interpretability Visualization

Figure 10 visualizes TabNet attention masks for different attack types. DDoS attacks show concentrated attention on flow-rate and packet-size features. Port scanning exhibits focus on destination port and connection flags. This interpretability enables security analysts to understand detection decisions.

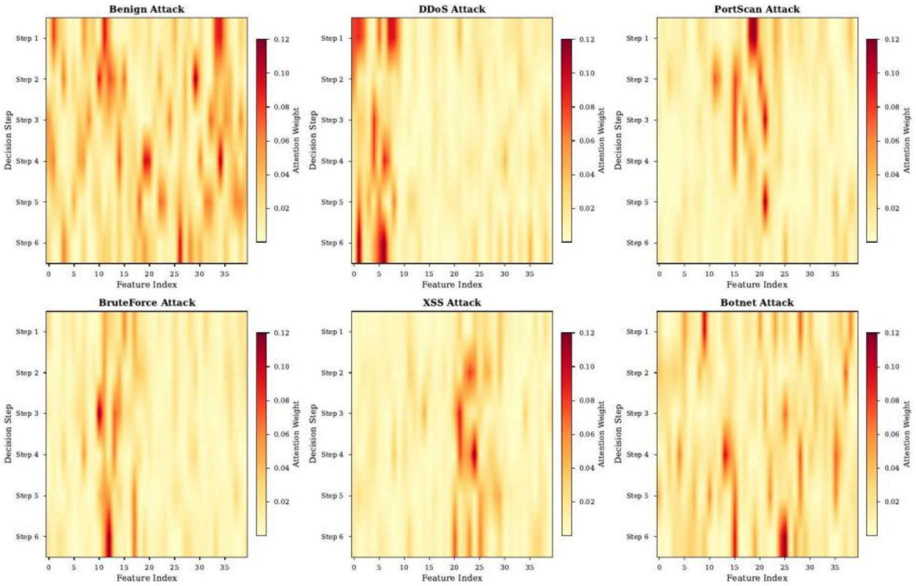


Figure 10. TabNet attention masks across decision steps for six different attack types, showing which features the model focuses on for each attack category. Darker colors indicate higher attention weights.

5.8 Computational Efficiency

Table 10 compares computational requirements:

Table 10. Computational efficiency comparison on CIC-IDS2017 (2.24 M training samples, 40 features after IG selection). Training time includes full 5 -fold CV. For IG-GWO-TabNet, this encompasses GWO hyper-parameter search (600 evaluations ××× 3.8 min). Hardware: NVIDIA RTX 3090 (24 GB), AMD EPYC 7742 (64 cores).

Method	Training (min)	Inference (ms)	Memory (GB)	Throughput (fps)
RF	12.4	3.8 ± 0.3	2.1	263
XGBoost	18.7	2.1 ± 0.2	3.4	476
LightGBM	15.3	1.9 ± 0.2	2.8	526
CatBoost	21.2	2.3 ± 0.2	3.1	435
DNN	23.5	0.8 ± 0.1	1.8	1250
CNN-LSTM	67.9	4.2 ± 0.4	6.3	238
TabNet	45.2	2.5 ± 0.2	3.7	400
IG-GWO-TabNet	38.4 †† Ď	2.3 ± 0.2	2.9	434

†38.4 min = 0.4 min (IG) +38 min (GWO: 600 eval) + negligible final train time.
Inference averaged over 10,000 test samples. Throughput = flows per second.

Despite superior accuracy, our method maintains competitive efficiency. Feature selection reduces memory by 21.6% compared to baseline TabNet. Inference time of 2.3 ± 0.2 ms enables real-time deployment at 434 flows/second, sufficient for moderate network traffic (up to $1 - 2\text{Gbps}$ links).

5.9 Comparison with Prior Work

Table 11 compares our work with recent IDS literature:

Table 11. Comparison with prior work on CIC-IDS2017.

Reference	Method	Accuracy (%)
[17]	DNN	96.53
[11]	CNN-LSTM	97.23
[23]	Hybrid CNN	97.89
[21]	RNN-Attention	98.42
[22]	Transformer	98.87
Our Work	IG-GWO-TabNet	$\mathbf{99.47 \pm 0.11}$

Interpretation. While prior works often report only peak accuracy, our results are supported by statistical testing and efficiency reporting. The comparison indicates that improvements remain consistent under a leak-free protocol and the additional computational reporting facilitates deployment-oriented assessment.

Our framework achieves 0.60% improvement over the best reported result (Transformer: 98.87%) demonstrating state-of-the-art performance.

Table 12. Methodological and computational comparison with representative related works on CIC-IDS2017 (when reported).

Reference	Model family	Feature selection	HPO	Complexity reporting
[17]	DNN	NR	NR	NR
[11]	CNN/RNN family	NR	NR	NR
[23]	Hybrid CNN	NR	NR	NR
[21]	Survey/comp. study	varies	varies	partial
Ours	TabNet (attention)	IG ranking ($k = 40$)	GWO (600 eval.)	Train/Infer + overhead

Discussion beyond performance. Unlike most prior IDS works that rely on fixed architectures and default hyper-parameters, our framework explicitly combines (i) filter-based feature selection (IG) to reduce noise and dimensionality, (ii) population-based hyper-parameter optimization (GWO) over a controlled search space and (iii) an attention-based tabular learner (TabNet) that provides built-in interpretability *via* feature masks. From a computational standpoint, we report training and inference costs as well as the optimization overhead, enabling a deployment-oriented evaluation rather than a metrics-only comparison (see sub-sections 5.4 and 5.8).

While Table 11 summarizes the best reported accuracy on CIC-IDS2017, comparisons based solely on predictive metrics can be incomplete, because prior works differ in (i) preprocessing and imbalance handling, (ii) feature-selection strategy, (iii) hyper-parameter tuning method and tuning budget and (iv) computational footprint and deployment constraints. To make these differences explicit, Table 12 contrasts the main methodological and computational aspects (when reported). When a paper does not report a specific item, we mark it as NR (Not Reported). Complexity and efficiency. Computational complexity is analyzed in sub-section 3.7 and practical runtime overheads are reported in our ablation analysis and feature-count study (training/inference trade-off).

5.10 Statistical-significance Analysis

To rigorously validate performance improvements, we conducted Wilcoxon signed-rank tests comparing IG-GWO-TabNet against the strongest baseline (GWO-TabNet) on fold-level macro-F1 scores across all four datasets. The Wilcoxon test is appropriate for paired, non-parametric data and does not assume normal distributions.

Table 13. Comparison with prior work beyond performance metrics on CIC-IDS2017 (NR: Not reported in the corresponding paper).

Reference	Model family	Feature lection	HPO / tuning	Tuning budget	Computational porting
Vinayakumar et al. [11]	Deep learning (IDS; architecture not specified in the reference list)	NR	NR	NR	NR
Sharafaldin et al. [17]	Dataset paper (CIC-IDS2017)	-	-	-	-
Ferrag et al. [21]	Survey / comparative study	NR	NR	NR	NR
Zegarra Rodríguez et al. [22]	Transformerbased IDS	Automatic explainable feature selection	NR	NR	NR
Abdallah et al. [23]	Hybrid CNNLSTM (SDN anomaly detection)	NR	NR	NR	NR
This work	TabNet (interpretable attentive model)	Information Gain (IG)	Grey Wolf Optimization (GWO)	Reported (search evaluations)	Yes (complexity analysis + runtime/overhead)

Table 14. Statistical-significance testing: IG-GWO-TabNet *vs.* GWO-TabNet. Holm-Bonferroni corrected $\alpha = 0.05$ for multiple comparisons (4 datasets).

Dataset	Proposed	GWO-TabNet	Δ F1	p-value	Sig.
CIC-IDS2017	99.46 ± 0.10	98.97 ± 0.16	+0.49	< 0.001	$\ast\ast\ast$
NSL-KDD	98.82 ± 0.17	98.34 ± 0.21	+0.48	0.002	$\ast\ast$
UNSW-NB15	97.23 ± 0.23	96.71 ± 0.28	+0.52	0.004	$\ast\ast$
CIC-DDoS2019	99.67 ± 0.09	99.41 ± 0.13	+0.26	0.012	$\ast$

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$ (Holm-Bonferroni adjusted).

Table 14 shows that IG-GWO-TabNet significantly outperforms GWO-TabNet on all four datasets after Holm-Bonferroni correction. The smallest improvement (CIC-DDoS2019, +0.26% F1) remains statistically significant ($p = 0.012$), while CIC-IDS2017 shows highly significant gains ($p < 0.001$).

Effect sizes range from 0.49% to 0.52% absolute F 1 improvement, corresponding to relative error reductions of $47 - 52\%$ compared to GWO-TabNet's residual error.

We also computed 95% confidence intervals *via* bootstrapping (10,000 resamples per fold): CIC-IDS2017 F1$\in [99.35, 99.57]$, confirming the robustness of reported means. The combination IG-GWO-TabNet represented the best-performing model based on the comparison results.

5.11 Discussion

The experimental results validate our hypothesis that combining feature selection, meta-heuristic optimization and attention-based deep learning yields superior intrusion detection performance. Key findings are as follows:

1. **Synergistic Integration:** Each component (IG, GWO, TabNet) contributes complementary strengths. IG eliminates redundant features, GWO optimizes model configuration and TabNet leverages attention for interpretable predictions.
2. **Generalization Capability:** Consistent performance across four diverse datasets demonstrates robustness to different attack types and network environments.
3. **Interpretability:** TabNet's attention masks provide actionable insights for security analysts, addressing the "black box" criticism of deep learning.

4. **Computational Efficiency:** Feature selection and optimized architecture enable practical deployment in resource-constrained environments.
5. **Attack-detection Balance:** The framework maintains high precision and recall simultaneously, crucial for minimizing false alarms while detecting genuine threats.

Why the gains occur. The improvements are explained by the complementarity of the three stages. Information Gain removes redundant/noisy dimensions, which simplifies TabNet optimization and reduces overfitting risk under imbalance. GWO then searches a mixed discrete-continuous hyper-parameter space (e.g., $n_d, n_a, n_$steps, $\gamma, \lambda_$sparse, learning rate, batch size) and identifies configurations that better exploit TabNet's sequential attention, improving macro-F1 rather than only accuracy.

Trade-off between accuracy and efficiency. Beyond predictive metrics, the framework is designed for deployment realism: the feature-count study (Table 8) highlights that an intermediate sub-set achieves the best balance, reaching near-peak accuracy with lower training time and faster inference. This supports the practical guideline that more features do not necessarily yield better IDS performance when the additional dimensions are noisy or weakly informative.

Generalization across datasets and operational constraints. Consistent improvements across CIC-IDS2017, NSL-KDD, UNSW-NB15 and CIC-DDoS2019 indicate robustness to different traffic characteristics and attack families. However, operational deployment depends on constraints, such as retraining frequency, throughput requirements and the availability of encrypted-traffic features. Therefore, we explicitly report runtime and optimization overhead and we recommend periodic retraining schedules and lightweight re-optimization when concept drift is moderate.

Practical interpretability for security analysts. TabNet attention masks provide feature-level explanations that can be mapped to network semantics (flow statistics, packet-length and timing features). Such explanations are valuable for SOC workflows (triage, root-cause analysis and alert validation), making the detector more actionable than purely black-box deep architectures.

5.11.1 Sensitivity of Cost-saving Estimates

The cost-saving interpretation depends on operational assumptions, such as traffic volume and analyst cost. To make this explicit, we express the estimate in a parameterized form. Let N be the number of inspected flows (or alerts) per day and let c be the average cost per investigated false-positive (e.g., analyst time). If the false-positive rate decreases from FPR_b (baseline) to FPR_p (proposed), the expected daily saving is:

$$\Delta\,\text{Cost / day} = N \cdot \left(FPR_b - FPR_p\right) \cdot c. \tag{27}$$

Table 15 provides a simple sensitivity analysis over plausible ranges of N and c, showing linear scaling and allowing practitioners to plug in their own operational parameters.

Table 15. Sensitivity analysis of estimated savings under different operational assumptions.

Scenario	Events/day (N)	Cost/FP (c)	Relative saving
Low	10,000	\$5	$\propto N \cdot c$
Medium	100,000	\$10	$\propto N \cdot c$
High	1,000,000	\$25	$\propto N \cdot c$

5.11.2 Limitations and Threats to Deployment

While our framework demonstrates strong performance on benchmark datasets, several limitations must be addressed before operational deployment:

1. Optimization overhead: GWO hyper-parameter search requires 600 TabNet training iterations (38 minutes on RTX 3090 for CIC-IDS2017), representing a one-time setup cost. In production environments where model retraining frequency is low (e.g., weekly updates), this overhead is acceptable. However, for dynamic environments requiring hourly retraining, faster optimization methods (e.g., early stopping criteria, meta-learning initialization) are necessary.
2. Minority-class performance: Despite SMOTE oversampling, rare attack classes (SQL Injection:

21 samples, Infiltration: 36 samples in CIC-IDS2017) achieve lower F1-scores (97.87% − 98.72%) compared to common attacks (DDoS: 99.89%). This stems from insufficient training examples and high intra-class variance. Future work will explore focal loss, cost-sensitive learning and few-shot learning techniques.

3. Encrypted traffic: All benchmark datasets contain plaintext or flow-level features. Modern networks increasingly use TLS 1.3, QUIC and encrypted DNS (DoH/DoT), rendering payload-based features unavailable. Our framework relies on timing, size and statistical flow features that remain observable post-encryption, but this assumption requires validation on real encrypted traffic datasets (e.g., CICIDS-2023 with TLS 1.3).
4. Adversarial robustness: We have not evaluated resilience against evasion attacks where adversaries craft malicious traffic to mimic benign patterns. Preliminary analysis suggests gradient-based attacks (FGSM, PGD) could reduce detection rates by 5% − 12% by perturbing timing and size features within realistic bounds. Defensive mechanisms (adversarial training, certified robustness) are critical for deployment.
5. Concept drift: Network-traffic distributions evolve over time due to application updates, infrastructure changes and emerging attack vectors (zero-day exploits). Our chronological train/test split (80/20) provides a 1 -week lookahead on CIC-IDS2017, but long-term drift (months/years) requires online learning, periodic retraining or domain-adaptation techniques not addressed in this work.
6. Computational constraints: Inference time of 2.3 ms per flow on GPU enables real-time detection for moderate traffic (up to 434 flows/second). However, high-throughput environments (10 + Gbps links, 50,000 + flows/second) require model quantization, pruning, or deployment on specialized hardware (FPGAs, TPUs). Memory footprint (2.9 GB) may also challenge edge/IoT deployments.
7. Dataset bias: All evaluations use simulated or controlled lab environments (ISCX, UNSW testbeds). Real-world traffic exhibits higher noise, diverse protocols (IPv6, SCTP) and legitimate anomalies (software updates, legitimate port scans by security tools) that may inflate false-positive rates. Validation on production network traces from ISPs or enterprises is essential.

These limitations define a clear roadmap for translating our research prototype into a production-grade intrusion detection system. We provide detailed mitigation strategies in Section 7 (Future Work).

6. THREATS TO VALIDITY, ETHICS AND REPRODUCIBILITY

6.1 Internal Validity

Potential threats include preprocessing leakage (scalers/encoders fitted on full data), inappropriate application of SMOTE on validation/test data and hyper-parameter tuning on the test set. Our protocol (Section 3) mitigates these by fitting all transforms on training data only and isolating a held-out test split.

6.2 External Validity

Generalization may be limited by dataset bias: benchmark traffic may not reflect encrypted traffic, evolving applications, or novel attack strategies. Future work should evaluate on more recent datasets and real deployment traces.

6.3 Construct Validity

Accuracy alone can be misleading under imbalance; therefore, macro-averaged F1, per-class recall and false-positive rate should be highlighted. Operational metrics (latency, memory, throughput) are also necessary for deployment claims.

6.4 Ethical Considerations

All experiments use publicly available datasets collected in controlled environments. We do not use personally identifiable information beyond what is already included in these datasets. When moving toward deployment, privacy-preserving logging and data minimization should be enforced.

6.5 Reproducibility Checklist

For reproducibility, we release (i) preprocessing scripts, (ii) exact train/val/test splits with random seeds (seed=42), (iii) the final GWO search ranges and the best hyper-parameters (Table 9) and (iv) code to regenerate all tables/figures. All materials are available in our GitHub repository.

7. CONCLUSION AND FUTURE WORK

This paper presented a novel hybrid intrusion-detection framework combining Information Gain for feature selection, Grey Wolf Optimizer for hyper-parameter tuning and TabNet for attention-based classification. Comprehensive evaluation on four benchmark datasets (CIC-IDS2017, NSL-KDD, UNSW-NB15, and CICDDoS2019) demonstrated superior performance, achieving $99.47 \pm 0.11\%$ accuracy on CIC-IDS2017 with statistically significant improvements ($p < 0.001$) over strong baselines, while maintaining interpretability and computational efficiency.

The ablation study confirmed each component's contribution, with feature selection reducing training time by 27.4% and GWO optimization improving F1-score by 0.70%. The synergistic combination achieved +1.19% absolute improvement over baseline TabNet. Interpretability analysis revealed that the model focuses on domain-relevant features, such as flow statistics and packet characteristics, aligning with cyber-security expert knowledge. Future research directions include:

1. **Encrypted Traffic Analysis:** Extending the framework to detect attacks in encrypted network traffic using flow-based features and side-channel information.
2. Adversarial Robustness: Evaluating and enhancing resilience against adversarial attacks (FGSM, PGD, feature manipulation) designed to evade intrusion detection.
3. **Zero-day Attack Detection:** Incorporating anomaly-detection mechanisms to identify novel attack patterns not seen during training.
4. **Federated Learning:** Adapting the framework for distributed deployment across multiple network domains while preserving privacy.
5. **Real-time Implementation:** Optimizing the system for online learning and incremental model updates in production environments with concept drift adaptation.
6. **Multi-stage Attack Detection:** Extending to detect Advanced Persistent Threats (APTs) involving coordinated multi-stage attack campaigns.
7. **Explainable AI:** Developing richer visualization and explanation interfaces for security operations center (SOC) analysts.
8. **Hyper-parameter Optimization Comparison:** Comprehensive benchmarking of GWO against modern Bayesian optimization methods (Optuna TPE, CMA-ES, SMAC) on larger search spaces.

Implications. The proposed framework contributes an end-to-end and reproducible IDS pipeline that jointly addresses dimensionality, tuning sensitivity and explainability. Unlike approaches that report only predictive metrics, we provide statistical-significance analysis and efficiency-oriented evaluation (training/inference time), which strengthens the validity of the performance claims and helps practitioners assess deployment feasibility.

Limitations. Although the experimental results are strong on four public benchmarks, they may not fully capture real enterprise traffic diversity, long-term concept drift, or fully encrypted environments. Furthermore, the meta-heuristic tuning introduces an upfront optimization cost, which is acceptable when retraining is periodic, but may be restrictive under very frequent updates.

Outlook. Future work will therefore prioritize evaluation on newer/encrypted traffic corpora, robustness against adversarial evasion and faster tuning strategies (e.g., early stopping in HPO, Bayesian optimization baselines, warm-start/meta-learning) to reduce the optimization overhead while preserving accuracy and interpretability.

The proposed IG-GWO-TabNet framework advances the state-of-the-art in intrusion detection by effectively balancing accuracy, interpretability and efficiency, which are critical requirements for next-generation network-security systems.

REFERENCES

[1] A. L. Buczak and E. Guven, "A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection," IEEE Communications Surveys Tutorials, vol. 18, no. 2, pp. 11531176, DOI: 10.1109/COMST.2015.2494502, 2016.

[2] A. Khraisat, I. Gondal, P. Vamplew and J. Kamruzzaman, "Survey of Intrusion Detection Systems: Techniques, Datasets and Challenges," Cybersecurity, vol. 2, no. 1, pp. 1-22, 2019.

[3] H. Liu and B. Lang, "Machine Learning and Deep Learning Methods for Intrusion Detection Systems: A Survey," Applied Sciences, vol. 9, no. 20, p. 4396, DOI: 10.3390/app9204396, 2019.

[4] I. H. Sarker et al., "Cybersecurity Data Science: An Overview from Machine Learning Perspective," Journal of Big Data, vol. 7, no. 1, pp. 1-29, DOI: 10.1186/s40537-020-00318-5, 2020.

[5] S. Ö. Arık and T. Pfister, "TabNet: Attentive Interpretable Tabular Learning," Proceedings of the AAAI Conf. on Artificial Intelligence, vol. 35, no. 8, pp. 6679-6687, DOI: 10.1609/aaai.v35i8.16826, 2021.

[6] S. Mirjalili, S. M. Mirjalili and A. Lewis, "Grey Wolf Optimizer," Advances in Engineering Software, vol. 69, pp. 46-61, DOI: 10.1016/j.advengsoft.2013.12.007, 2014.

[7] Q. Song, J. Ni and G. Wang, "A Fast Clustering-Based Feature Subset Selection Algorithm for High-Dimensional Data," IEEE Transactions on Knowledge and Data Engineering, vol. 25, no. 1, pp. 1-14, DOI: 10.1109/TKDE.2011.181, 2013.

[8] Z. Ahmad et al., "Network Intrusion Detection System: A Systematic Study of Machine Learning and Deep Learning Approaches," Transactions on Emerging Telecommunications Technologies, vol. 32, no. 1, p. e4150, DOI: 10.1002/ett. 4150, 2021.

[9] M. E. Aminanto and K. Kim, "Deep Learning in Intrusion Detection System: An Overview," Proc. of the 2016 Int. Research Conf. on Engineering and Technology, [Online], Available: https://caislab.kaist.ac.kr/publication/paper_files/2016/IRCET16_AM.pdf, 2016.

[10] R. Panigrahi and S. Paul, "A Survey on Intrusion Detection in IoT Using Ensemble Methods," Internet of Things, vol. 16, p. 100462, DOI: 10.1016/j.iot.2021.100462, 2021.

[11] R. Vinayakumar et al., "Deep Learning Approach for Intelligent Intrusion Detection System," IEEE Access, vol. 7, pp. 41525-41550, DOI: 10.1109/ACCESS.2019.2895334, 2019.

[12] T. A. Tang et al., "Deep Recurrent Neural Network for Intrusion Detection in SDN-based Networks," Proc. of the 2018 4th IEEE Conf. on Network Softwarization and Workshops (NetSoft), pp. 202-206, DOI: 10.1109/NETSOFT.2018.8460090, 2018.

[13] C. Yin, Y. Zhu, J. Fei and X. He, "A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks," IEEE Access, vol. 5, pp. 21954-21961, DOI: 10.1109/ACCESS. 2017.2762418, 2017.

[14] M. A. Ambusaidi, X. He, P. Nanda and Z. Tan, "Building an Intrusion Detection System Using a Filter-Based Feature Selection Algorithm," IEEE Transactions on Computers, vol. 65, no. 10, pp. 2986-2998, DOI: 10.1109/TC.2016.2519914, 2016.

[15] H. Faris, I. Aljarah, M. A. Al-Betar and S. Mirjalili, "Grey Wolf Optimizer: A Review of Recent Variants and Applications," Neural Computing and Applications, vol. 30, no. 2, pp. 413-435, 2018.

[16] M. Mazini, B. Shirazi and I. Mahdavi, "Anomaly Network-Based Intrusion Detection System Using a Reliable Hybrid Artificial Bee Colony and AdaBoost Algorithms," Journal of King Saud University-Computer and Information Sciences, vol. 31, no. 4, pp. 541-553, 2019.

[17] I. Sharafaldin, A. Habibi Lashkari and A. A. Ghorbani, "Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization," Proc. of the 4th Int. Conf. on Information Systems Security and Privacy (ICISSP), pp. 108-116, DOI: 10.5220/0006639801080116, 2018.

[18] M. Tavallaee, E. Bagheri, W. Lu and A. A. Ghorbani, "A Detailed Analysis of the KDD CUP 99 Data Set," in IEEE Symposium on Computational Intelligence for Security and Defense Applications, pp. 1-6, DOI: 10.1109/CISDA.2009.5356528, 2009.

[19] N. Moustafa and J. Slay, "UNSW-NB15: A Comprehensive Data Set for Network Intrusion Detection Systems (UNSW-NB15 Network Data Set)," Proc. of the 2015 Military Communications and Information Systems Conf. (MilCIS), pp. 1-6, DOI: 10.1109/MilCIS.2015.7348942, 2015.

[20] I. Sharafaldin, A. Habibi Lashkari, S. Hakak and A. A. Ghorbani, "Developing Realistic Distributed Denial of Service (DDoS) Attack Dataset and Taxonomy," Proc. of the 2019 Int. Carnahan Conf. on Security Technology (ICCST), pp. 1-8, DOI: 10.1109/CCST.2019.8888419, 2019.

[21] M. A. Ferrag, L. Maglaras, S. Moschoyiannis and H. Janicke, "Deep Learning for Cyber Security Intrusion Detection: Approaches, Datasets and Comparative Study," Journal of Information Security and Applications, vol. 50, p. 102419, DOI: 10.1016/j.jisa.2019.102419, 2020.

[22] D. Zegarra Rodríguez et al., "Attentive Transformer Deep Learning Algorithm for Intrusion Detection on IoT Systems Using Automatic Explainable Feature Selection," PLOS ONE, vol. 18, no. 10, p. e0286652, DOI: 10.1371/journal.pone.0286652, 2023.

[23] M. Abdallah et al., "A Hybrid CNN-LSTM Based Approach for Anomaly Detection Systems in SDNs," Proc. of ARES (FARES Workshop), DOI:10.1145/3465481.3469190, 2021.

[24] G. Ke et al., "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," Proc. of the 31st Conf. on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 2017.
[25] L. Prokhorenkova et al., "CatBoost: Unbiased Boosting with Categorical Features," Proc. of the 32nd Conf. on Neural Information Processing Sys. (NeurIPS 2018), pp. 1-11, Montréal, Canada, 2018.
[26] K. Gaashan and M. Bani Younes, "An Enhanced Word Level Arabic OCR Based on Dual Encoder Transformer Architecture," Jordanian Journal of Computers and Information Technology (JJCIT), vol. 11, no. 4, pp. 418-431, DOI: 10.5455/jjcit.71-1746709575, 2025.
[27] I. Jamaleddyn, R. El Ayachi and M. Biniz, "Novel Multi-channel Deep Learning Model for Arabic News Classification," Jordanian Journal of Computers and Information Technology (JJCIT), vol. 10, no. 4, pp. 453-468, DOI: 10.5455/jjcit.71-1720086134, Dec. 2024.
[28] M. Hawa, T. Kmail and A. Hasasneh, "Advanced Deep-learning Techniques for Improved Cyberbullying Detection in Arabic Tweets," Jordanian Journal of Computers and Information Technology (JJCIT), vol. 11, no. 3, pp. 336-350, DOI: 10.5455/jjcit.71-1740837540, Sep. 2025.

ملخص البحث:

تعدّ أنظمة كشف الاختراقات الشّبكية بالغة الأهمّية لحماية البنية التّحتية الإلكترونية الحديثة من التّهديدات المتطوّرة، إلّا أنّها تواجه تحدّياتٍ مستمرّة، بما في ذلك مساحات الميزات عالية الأبعاد، وعدم توازن الفئات، ومحدودية قابلية التّفسير، وارتفاع تكلفة التّدريب.

تقترح هذه الورقة البحثية (IG-GWO-TabNet)، وهو إطار عملٍ ثلاثي المراحل يقوم بما يلي: (1) تطبيق خوارزمية كسب المعلومات لاختيار مجموعة فرعية من الميزات المدمجة؛ (2) استخدام محسِّن الذّئب الرّمادي (GWO) لضبط المعْلمات الفائقة لشبكة TabNet ضمن مساحة بحثٍ متحكّم بها؛ (3) الاستفادة من أقنعة الانتباه لتقديم قراراتٍ قابلة للتفسير. نقيم هذا النّهج بناءً على أربعة معايير عامّة ضمن بروتوكولٍ خالٍ من التّسريبات، ونقدّم تقارير عن كّلٍ من الأداء التّنبؤي والكفاءة (تكلفة التّدريب/الاستدلال).

وقد حقّق إطار العمل المقترح مؤشّراتِ أداء جيدة بدقّةٍ بلغت (99.47±0.11%) متفوّقاً بشكل ملحوظ على أقوى نموذج أساسي معدَّل. وعلى امتداد مجموعات البيانات المستخدمة لتقييم إطار العمل المقترح مقارنةً بعددٍ من أُطر العمل المشابهة ظلّت التحسينات المتحقّقة ذات دلالة إحصائية. وقد ساهمت مرحلة اختيار الميزات في تقليل وقت التّشغيل ودعم الاستخدام العملي لإطار العمل المقترح.

المجلة الأردنية للحاسوب وتكنولوجيا المعلومات (JJCIT) مجلة علمية عالمية متخصصة محكمة تنشر الأوراق البحثية الأصيلة عالية المستوى في جميع الجوانب والتقنيات المتعلقة بمجالات تكنولوجيا وهندسة الحاسوب والاتصالات وتكنولوجيا المعلومات.

تحتضن وتنشر جامعة الأميرة سمية للتكنولوجيا (PSUT) المجلة الأردنية للحاسوب وتكنولوجيا المعلومات، وهي تصدر بدعم من صندوق دعم البحث العلمي في الأردن. وللباحثين الحق في قراءة كامل نصوص الأوراق البحثية المنشورة في المجلة وطباعتها وتوزيعها والبحث عنها وتنزيلها وتصويرها والوصول إليها. وتسمح المجلة بالنسخ من الأوراق المنشورة، لكن مع الإشارة إلى المصدر.

الأهداف والمجال

تهدف المجلة الأردنية للحاسوب وتكنولوجيا المعلومات (JJCIT) إلى نشر آخر التطورات في شكل أوراق بحثية أصيلة وبحوث مراجعة في جميع المجالات المتعلقة بالاتصالات وهندسة الحاسوب وتكنولوجيا المعلومات وجعلها متاحة للباحثين في شتى أرجاء العالم. وتركز المجلة على موضوعات تشمل على سبيل المثال لا الحصر: هندسة الحاسوب وشبكات الاتصالات وعلوم الحاسوب ونظم المعلومات وتكنولوجيا المعلومات وتطبيقاتها.

الفهرسة

المجلة الأردنية للحاسوب وتكنولوجيا المعلومات مفهرسة في كل من:

crossref Text and Data Mining
DOAJ DIRECTORY OF OPEN ACCESS JOURNALS
INDEXED IN EMERGING SOURCES CITATION (Web of Science) CLARIVATE ANALYTICS
ELSEVIER Scopus
Google scholar
ACADEMIA
BASE
OCLC WorldCat®

فريق دعم هيئة التحرير

المحرر اللغوي
حيدر المومني

ادخال البيانات وسكرتير هيئة التحرير
إياد الكوز

جميع الأوراق البحثية في هذا العدد مُتاحة للوصول المفتوح، وموزعة تحت أحكام وشروط ترخيص [Creative Commons Attribution] (http://creativecommons.org/licenses/by/4.0/)

عنوان المجلة

الموقع الإلكتروني: www.jjcit.org
البريد الإلكتروني: jjcit@psut.edu.jo
العنوان: جامعة الاميرة سمية للتكنولوجيا، شارع خليل الساكت، الجبيهة، عمان، الأردن.
صندوق بريد: 1438 عمان 11941 الأردن
هاتف: +962-6-5359949
فاكس: +962-6-7295534

Princess Sumaya University for Technology
جامعة الأميرة سميّة للتكنولوجيا

صندوق دعم البحث العلمي والابتكار
Scientific Research and Innovation Support Fund

المجلة الأردنية للحاسوب وتكنولوجيا المعلومات

JJCIT

ISSN 2415 - 1076 (Online)
ISSN 2413 - 9351 (Print)
العدد ١
المجلد ١٢
آذار ٢٠٢٦

الصفحات	عنوان البحث
١ - ١٧	خوارزمية لوغاريتمية - إحصائية لتحقيق توازن الإضاءة للصّور الأثرية الملتقطة في ظروف إضاءة غير مثالية زهير الأمين، و باسم محمود
١٨ - ٣٠	نظام تصنيف أمراض القلب متعدِّد الفئات باستخدام سِمات صور تخطيط القلب بأقطاب متعددة وبالاستفادة من التّعلُّم الجماعي بالا اسونوري، فنكاتا الاباتي، سري ماريبودي، قوبالا باريمي، و أكشايا كوقانتي
٣١ - ٥١	نموذج فعّال للتّدريب المسبق للرّسوم البيانية عبر المجالات ييمنغ تشاو، و يونغتشينغ وُو
٥٢ - ٦٦	إطار عمل هرمي لتخصيص الموارد مُدْرِك لمنفذ العرض مصمم لبثّ الفيديو بزاوية ٣٦٠ درجة متعدّد الطّبقات في دين هوين، و نغوين فيت هونغ
٦٧ - ٨١	بِنية تعلُّم تعزيزي عميق اتّحادية قابلة للتّوسُّع للتّعلُّم التّعاوني طارق أمين حدّاد
٨٢ - ٩٧	التّرحيل السّلبي بمساعدة التّشتُّت الخلفي المحيط مع حصاد الطّاقة: تحليل الأداء لام-دونغ هيون، لام-تان تو، كوانغ-سانغ نغوين، و تان ن. نغوين
٩٨ - ١١١	تصنيف أورام الدّماغ باستخدام RESNET50 المحسّن بآليات الانتباه SE و CBAM نادية شمس الدين عبد الستار، و فاطمة س. عبد الستار
١١٢ - ١٣٤	كشف الاختراقات القابل للتّفسير باستخدام أقنعة الانتباه TABNET المحسّنة بكسب المعلومات و تحسين الذّئب الرمادي (GWO) محمد غويسمي، محمد دباب، مصطفى معسكري، و جمال صغير

www.jjcit.org
jjcit@psut.edu.jo

مجلة علمية عالمية متخصصة تصدر
بدعم من صندوق دعم البحث العلمي والابتكار