



جامعة
الأميرة سميرة
للتكنولوجيا
Princess Sumaya
University
for Technology



صندوق دعم البحث العلمي والابتكار
Scientific Research and Innovation Support Fund

Jordanian Journal of Computers and Information Technology

March 2024

VOLUME 10

NUMBER 01

ISSN 2415 - 1076 (Online)
ISSN 2413 - 9351 (Print)

PAGES

PAPERS

1 - 16	MULTI-DOMAIN MACHINE LEARNING APPROACH OF NAMED ENTITY RECOGNITION FOR ARABIC BOOKING CHATBOT ENGINES USING PRE-TRAINED BIDIRECTIONAL TRANSFORMERS Boshra Sadder, Rahma Sadder, Gheith Abandah and Iyad Jafar
17 - 31	RADIO FREQUENCY IDENTIFICATION NETWORK PLANNING USING ASCENDANT HIERARCHICAL CLUSTERING ALGORITHM Lamri Sayad
32 - 45	TO DISTRIBUTE THE ABNORMAL EVENT USING CM-DDS ALGORITHM, WITH SECURELY USING RSA-IVDE ALGORITHM Kande Archana and V. Kamakshi Prasad
46 - 57	ARABIC SOFT SPELLING CORRECTION WITH T ₅ Mohammed Al-Qaraghuli and Ola Arif Jaafar
58 - 73	KOCH FRACTAL OCTAGONAL ANTENNA WITH A COMPACT DESIGN AND DEFECTED GROUND STRUCTURE (DGS) FOR ULTRA-WIDEBAND (UWB) WIRELESS USAGE Tejaswita Kumari, Abu Nasar Ghazali and Anupama Senapati
74 - 92	CDRSHNET: VARIANCE GUIDED MULTISCALE AND SELF-ATTENTION FUSION WITH HYBRID LOSS FUNCTION TO RESTORE TRAFFIC SIGN IMAGE SCAPTURED IN ADVERSE CONDITIONS Milind Vijay Parse and Dhanya Pramod
93 - 107	USING RESNET ₁₈ IN A DEEP LEARNING FRAMEWORK AND ASSESSING THE EFFECTS OF ADAPTIVE LEARNING RATES IN THE IDENTIFICATION OF MALIGNANT BREAST MASSES IN MAMMOGRAMS Soumia BENBAKRETI, Samir BENBAKRETI, Kadda BENYAHIA and Mohamed BENOUS

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:



EDITORIAL BOARD SUPPORT TEAM

LANGUAGE EDITOR

Haydar Al-Momani

EDITORIAL BOARD SECRETARY

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	Nijad Al-Najdawi
Hussein Al-Majali	Maen Hammad	Ayman Abu Baker
Ahmed Al-Taani	João L. M. P. Monteiro	Leonel Sousa
Omar Al-Jarrah		

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".

MULTI-DOMAIN MACHINE LEARNING APPROACH OF NAMED ENTITY RECOGNITION FOR ARABIC BOOKING CHATBOT ENGINES USING PRE-TRAINED BIDIRECTIONAL TRANSFORMERS

Boshra Sadder, Rahma Sadder, Gheith Abandah and Iyad Jafar

(Received: 12-Sep.-2023, Revised: 4-Nov.-2023, Accepted: 20-Nov.-2023)

ABSTRACT

Chatbots have recently become essential in various fields, ranging from customer service and information acquisition to entertainment. The use of chatbots reduces operational costs and human errors while providing services at any time. This work presents a Named Entity Recognition (NER) model for the Arabic booking chatbot, focusing on booking tickets and appointments across multiple domains. This research paves the way for the development of chatbots that can support multiple booking domains, contributing to the advancement of the Arabic language in this field. We adopt deep machine-learning and transfer-learning approaches to solve this task. Specifically, we utilized and fine-tuned the AraBERTv0.2 base model to develop the Named Entity Recognition for Booking Queries (NERB) model. Furthermore, we extended it to the Domain-Aware Named Entity Recognition for Booking Queries (DA-NERB) model by adding an additional input for domain type and an embedding layer. The input to our proposed model consists of text sequences of reservation requests, while the output includes sequences of tags representing entities within the input sequences. For training and testing, we synthesized the Arabic Booking Chatbot-Synthetic Dataset (ABC-S Dataset), comprising 76,117 reservation samples that span seven different domains and encompassing 26 categories of named entities. Additionally, we collected the Arabic Booking Chatbot-Collected Dataset (ABC-C Dataset) from volunteers to evaluate our model using various samples. It's worth noting that these datasets are written in informal Arabic, specifically the Levantine dialect. The proposed model achieves 100% and 96.9% accuracy scores on ABC-S (test set) and ABC-C, respectively. Both the datasets and the code for our model are publicly available to support research in the field of Arabic chatbots.

KEYWORDS

Chatbot, Arabic booking chatbot, Named entity recognition, AraBERT, Arabic booking dataset.

1. INTRODUCTION

Human-computer interaction is a technology that enables the interaction between humans and computers using natural languages, such as English and Arabic. The complexity of human language has necessitated a technology capable of understanding human language [1]. In the past decades, artificial intelligence (AI) and deep learning (DL) have shown significant development in various areas, including computer vision, natural-language processing (NLP) and speech processing [2]. NLP is a branch of AI research that explores computers' ability to understand human language.

A chatbot, also known as a conversation agent (CA), represents an example of an AI application that facilitates interaction between a person and a computer device using natural language. Chatbot agents have garnered the attention of numerous researchers aiming to make the conversation between humans and machines more rational [3].

Thanks to the significant advancements in AI and machine learning, users view chatbots as a promising alternative to traditional customer-service channels. They are perceived by customers as natural and function as virtual assistants, offering a more interactive experience compared to mobile applications. Chatbots help perform tasks, address customer inquiries and make purchase recommendations. Over the past few years, there has been a growing interest in chatbots. In general, chatbots offer instant responses, can serve an unlimited number of users simultaneously and provide 24-hour service [4]. This has a direct impact on businesses in terms of reducing the cost of delivering services to customers. In summary, chatbots have the potential to save money, manpower and possibly

increase customer satisfaction [5].

Building chatbots that support the English language has received more attention and is technically more advanced than those designed for Arabic. The scarcity of Arabic chatbots can be attributed to several factors. Primarily, these include: 1) The lack of training data in Arabic due to the widespread use of English in comparison to Arabic. 2) The complexity and characteristics of the Arabic language, including the fact that it is a Semitic language rich in morphology and featuring variations in orthography. 3) The presence of diverse Arabic dialects [6]-[7].

Numerous chatbots specialize in booking services, such as those for booking airline tickets, hotels, ...*etc.* However, most of the current research on modeling booking chatbots focuses on a single domain, with only a few supporting multiple domains. Furthermore, when it comes to the Arabic language, there is a shortage of efficient booking chatbots. In general, published studies on Arabic chatbots are scarce, highlighting a significant knowledge gap [7].

Chatbots expect users to submit booking queries in the form of unstructured text and they are supposed to process these queries and respond accordingly. One crucial step in this process is the chatbot's ability to recognize critical and important entities within the query to formulate an appropriate response. The main objective of this work is to develop a named-entity recognition (NER) model for Arabic booking chatbots, which can be applied across multiple domains to extract and classify reservation entities in booking queries with high accuracy. To build our model, we used and fine-tuned the state-of-the-art AraBERTv0.2 model.

A significant contribution of this work is the building of the Arabic Booking Chatbot-Synthetic Dataset (ABC-S Dataset), which contains 76,117 samples and encompasses 26 entity categories. Additionally, we collected reservation samples from volunteers to evaluate our model on various samples; this dataset is referred to as the Arabic Booking Chatbot-Collected Dataset (ABC-C Dataset). Also, we propose a solution for the Arabic booking chatbot that is dedicated to booking tickets and appointments for seven different domains, which are flights, hotels, cinemas, football matches, cars, restaurants and clinics. The main characteristic of the proposed model when compared to similar models lies in the ability to perform the NER task when different booking domains are considered. In other words, we present a single NER model that is domain-aware, eliminating the need for multiple models.

The rest of the paper is organized as follows, In Section 2, we provide a review of related work on chatbots and NER models. Section 3 discusses the process of building and collecting the datasets, as well as describing the two NER models. In Section 4, we present and discuss the evaluation results and outline the limitations of this work. Finally, we conclude the paper in Section 5.

2. LITERATURE REVIEW

There are many works on Arabic chatbot applications. The first part of the literature review presents papers on chatbot applications in Arabic. The second part of the literature review presents the papers that include NER models, as well as the main papers on which this research is based.

2.1 Chatbot Applications

Chatbot systems are basically software applications used to conduct a conversation between humans and computers using natural languages. These systems have recently become essential in our lives, as there are many organizations that use them to provide customer service.

The Dialogue System Development Framework is a set of services, tools and software-development kits that provide a rich foundation or framework for AI-chatbot development [8]. For example, it provides web interfaces to create a knowledge base. With the advent of these development frameworks, the implementation of the dialogue system has become easier and many dialogue systems (DSs) have been introduced to serve different tasks. Examples of such frameworks are: Wit.ai, Rasa and IBM Watson Assistant [9].

In the past few years, many chatbots have been developed for various languages to serve several domains. Unfortunately, there is limited research on Arabic chatbots due to the characteristics of the Arabic language. Al-Hammoud et al. [3] surveyed the available research conducted in the field of

Arabic chatbots and concluded that there is a scarcity of research available on Arabic chatbots and that all the available works are retrieval-based [3].

Al-Ajmi and Al-Twairesh [10] proposed an Arabic dialogue system for flight booking tasks using a hybrid rule-based and data-driven approach utilizing the Wit.ai framework, since it has a set of predefined trained entities that support Arabic [10]. This system is supposed to process the text provided by the user, understand it and then extract a set of entities, such as location, route type, date, time and ticket class. In addition, the system must be able to recognize whether the user's input contains an intention to book a flight to proceed with the reservation process.

The evaluation of the developed system was conducted holistically, without focusing on individual components. The assessment involved 21 participants, with 90.48% of them having prior experience in booking flight tickets. The evaluation of the Dialogue System (DS) took place in two stages, aiming to assess both ease of use and system effectiveness. In the initial stage, participants provided feedback to enhance the system, which informed improvements made to the developed models for testing in the subsequent stage. Overall, users had a positive experience when booking airline tickets using the developed flight-booking DS.

Al-Ghadhban and Al-Twairesh [11] developed an Arabic chatbot called "Nabiha," designed to serve as an academic advisor for students in the Information Technology Department at King Saud University (KSU). Nabiha's primary objective is to engage with students, providing answers to their inquiries regarding academic progress and available courses. The chatbot's development relied on pattern matching and the utilization of the Artificial Intelligence Markup Language (AIML). To convert readable text into AIML format, a Java program was created for this purpose. Data was collected from websites containing student opinions and complaints and the Nabiha chatbot was launched on the Pandorabots platform [11].

Fadhil [12] introduced OlloBot, an Arabic conversational agent that helps physicians follow-up with their patients and provide 24/7 support to patients. This chatbot utilized the IBM Watson Conversation platform to manage dialogue flow and provide AI assistance for capturing various user intents and entities. Following the development of the chatbot system, it was integrated with the Telegram Bot Platform [12]. To evaluate the system's performance, a total of forty-three users participated. The evaluation involved a questionnaire comprising 30 questions, focusing on four main aspects: usefulness, ease of use, ease of learning and user satisfaction with OlloBot's reliability. The respondents rated the questionnaire on a scale from 1 (strongly disagree) to 5 (strongly agree). The results showed good indicators, with most users expressing interest in the bot's social intelligence as well.

Alshareef and Siddiqui [6] presented an open-domain conversational agent (CA) designed to communicate in the Gulf dialect. They employed a sequence-to-sequence architecture model incorporating the attention mechanism to handle long dependencies in the input. The conversational problem was defined as a machine-translation problem. Therefore, they collected their corpus from tweets in the post-reply format. Their proposed model underwent training both with and without the use of pre-trained word embeddings and FastText. The system's evaluation included the Bilingual Evaluation Understudy (BLEU) score and human assessment. The results demonstrated that utilizing embeddings led to more comprehensible and contextually relevant outputs. Specifically, BLEU scores of 25.1 and 11.4 were achieved with and without the use of pre-trained word embeddings, respectively [6].

2.2 Named Entity Recognition (NER) Models

One major task of goal-oriented chatbot systems is to fill a set of slots embedded in a semantic frame or automatically extract essential information, including intents and entities, to accomplish the objective of the dialogue [13]. Solving this task is usually done using the NER technique, which aims to identify and categorize named entities within the user's input into predefined classes, such as person names, organizations and locations. Since the task is considered a multi-class classification process, some methods of text classification are used to name the tokens [14].

Devlin *et al.* [15] introduced a deep bidirectional-language model called BERT, which stands for bidirectional encoder representations from transformers. Language-specific BERT models are very effective in understanding language, provided that they are pre-trained on a very large corpus. The

pre-trained BERT model can be fine-tuned with just one additional output layer to efficiently solve many NLP tasks, such as question answering and named entity recognition (NER) [15].

BERT is a multi-layer bidirectional transformer encoder. The BERT-Base model consists of 12 layers, 768-hidden size, 12 self-attention heads and 110M parameters, while the BERT Large model consists of 24 layers, 1024-hidden size, 16 self-attention heads and 340M parameters [15]. This bidirectional model can read the text input once, unlike the unidirectional model, which reads the text input sequentially from one direction. This feature allows the BERT model to learn the context of the word in both directions [16]. In the pre-training stage of the BERT transformer, the model is trained using an unsupervised approach on a very large corpus of data over two tasks, the masked language models (MLM) task and the next-sentence prediction (NSP) task.

This transformer has a very large number of parameters that can be fine-tuned for a certain NLP task by adding an additional layer or more on top of BERT that fits the shape of the output and then training the entire model together. In this stage, the model parameters are first initialized with the pre-training parameters and then are fine-tuned using the supervised approach by training the model using a labeled dataset associated with the NLP task [15]. The BERT model expects a text input; however, this text must be segmented into words (tokens). In addition, some special tokens are entered with the input tokens, such as the classification token (CLS) which is used at the beginning of BERT inputs and the separation token (SEP) which is used to separate the parts of the input. The output of BERT is a vector representation for each word as the output [17].

Antoun *et al.* [14] proposed the AraBERT model by pre-training BERT on a large-scale Arabic corpus. AraBERT was evaluated on three different natural-language processing tasks; namely, sentiment analysis, question answering and NER using eight different datasets. The results of AraBERT's experiments are compared with Google's multilingual BERT (mBERT) and other state-of-the-art approaches. The performance of the AraBERT outperformed all approaches on most of the tested datasets [14]. There are different versions of the AraBERT, which are AraBERTv0.1, AraBERTv0.2, AraBERTv1 and AraBERTv2 with the main differences between them being the size of the model and the size of the used training datasets, in addition to the need for pre-segmentation of the input text before it is fed to the model.

For the NER task, Antoun *et al.* [18] fine-tuned the AraBERTv0.1 model to solve the NER task on the Arabic Named Entity Recognition Corpus (ANERcorp), which encompasses 16.5 k entities categorized into four groups: person, organization, location and miscellaneous. Their approach involved customizing the AraBERTv0.1 model by adding two new layers. The first was a dropout layer with a rate of 0.3, while the second was a linear output layer. The model was trained using Adam optimization, a learning-rate scheduler, cross-entropy loss function and 5-fold cross-validation. They reported an F1-score of 84.2% [18].

Youssef *et al.* [19] proposed an end-to-end deep learning NER model for Arabic. This model consists of stacked embeddings that combine the pooled contextual embedding, pre-trained word embeddings (FastText) and BERT embeddings (pre-trained AraBERT model), that are fed into bidirectional long short-term memory with a conditional random field layer (BiLSTM-CRF). The results showed that this model outperformed all previously published results for deep learning and non-deep learning models on the AQMAR dataset with an F1-score of 77.62% [19].

Taher *et al.* [20] proposed a model for NER in Persian. In the architecture of their model, they used the BERT pre-trained model, a fully connected layer and a conditional random field (CRF) layer. These two layers were trained on the representation of tokens that are extracted from BERT. In the NSURL-2019 competition for the NER task for the Persian language, this model achieved second place with F1-scores of 83.5% and 88.4%, in phrase and word level evaluation, respectively [20].

Benali *et al.* [21] presented the BERT-BiLSTM-CRF model, which incorporated the BERT model as an embedding feature and integrated it with the BiLSTM-CRF architecture. The outcomes of this study demonstrated its model's superiority over six state-of-the-art Arabic NER models, achieving the highest performance on a tweet corpus with an F1-score of 67.40% [21].

Shaker *et al.* [22] proposed the utilization of Long Short-Term Memory (LSTM) units and Gated Recurrent Units (GRUs) for building named entity recognition models in the Arabic language.

Furthermore, they introduced a diverse Arabic NER dataset encompassing nine categories of named entities, comprising over thirty-six thousand records, with texts spanning seven different domains. Both the LSTM and GRU models produced commendable results, achieving an approximate precision of 80% in identifying entity names [22].

NER models play a pivotal role in the extraction and classification of named entities, providing significant utility in various chatbot applications, particularly in the domain of booking services. Despite the extensive attention that NER has received in English, a notable gap persists in its comprehensive exploration within Arabic. This gap can be attributed to the morphological complexity of the language and the limited availability of Arabic resources [23]. Additionally, most of the prior research on Arabic conversational systems primarily focused on Modern Standard Arabic (MSA), with relatively few incorporating Arabic dialects, even though numerous dialects are spoken across the Arabic-speaking world [7]. Furthermore, publicly accessible Arabic datasets often lack support for multiple reservation domains. Hence, the objective of this work is to advance the field of Arabic NER by developing an accurate, domain-aware NER model and building a specialized dataset encompassing seven different reservation domains for training our model.

3. MATERIALS AND METHOD

3.1 The Developed Datasets

Training deep machine-learning models for NLP tasks typically requires thousands of samples to achieve satisfactory results. In our case, this means the need to gather a substantial number of booking samples for various domains and manually tag the words in these samples, which can be a challenging task. Alternatively, we have proposed the creation and synthesis of a dataset for Arabic booking queries automatically using programs. We refer to this dataset as the Arabic Booking Chatbot-Synthetic Dataset (ABC-S Dataset). Additionally, for a more realistic evaluation and results, we have also collected another dataset of real Arabic booking queries from volunteers, which we named the Arabic Booking Chatbot-Collected Dataset (ABC-C Dataset). These datasets are used to train and test the proposed models. Both datasets are publicly available on GitHub [24].

3.1.1 Building the ABC-S Dataset

The dataset to be designed is intended to support multi-domain NER in Arabic booking queries. There are many domains in which booking service is required. In the ABC-S dataset, we proposed populating it with booking queries in the flight, hotel, cinema, football match, car rental, restaurant and clinic domains.

Next, for each of these domains, we listed the possible entities that may be present in the queries. For example, when booking a flight ticket, the query may include the departure and arrival cities, departure and return dates and times, the number of tickets and the ticket class. To facilitate the learning process of the NER model, the identified entities were assigned numeric labels or tags. A total of 25 entities were defined for the seven domains. Table 1 lists the names of these entities and their corresponding tags. Note that some entities are common between different domains, such as "DateAndTime 1," while others are exclusive to one domain, like the "DoctorName" entity. The entity name 'Others' has been added to the list and given a tag of 0. This entity is used to label any entity in the booking query that is not among the 25 predefined entities.

With these entities and their defined tags, we created a set of various reservation templates for each domain. These templates are written in informal Arabic (Levantine dialect) and essentially consist of sentences for possible reservation queries that contain general text and the names of entities. Table 2 displays some of the booking templates in the flight domain. The entity names, represented by the colored underlined words in the template, serve as variables and are later filled with actual values, which are then concatenated with the remaining text when all templates are generated.

To generate complete reservation samples in the dataset, a list of values for each of these entities was defined. For instance, the "Flight Ticket Class" entity has five different values, as listed in the first column of Table 3. A list of tag sequences was created to label each word in the value with the appropriate entity tag. Table 3 illustrates how the various possible values for the "Ticket Class" entity are tagged, with each word assigned a tag value of 7, corresponding to the "Ticket Class" tag, as shown in Table 1.

testing. The training set was further divided into training and validation sets, with percentages of 90% and 10%, respectively. Each of these sets is stored in a separate file with the structure shown in Table 6. The first column contains the sentence used to make the reservation, the second column contains a sequence of tags representing the entities in the sentence, the third column contains the domain ID to which the sample belongs and the fourth column contains the sample's label in the format (domain ID-template ID).

Table 4. Steps for generating a reservation sample in the flights domain.

Step 1. Select a template	"أريد حجز رحلة ذهاب من " + City 1 + " إلى " + City 2 + " " + Numbers + " " + TicketClass + " على " + Time 1 + " الساعة " + DateAndDay 1 <i>"I want to book one-way ticket from " + City 1 + " to " + City 2 + " " + Numbers + " " + DateAndDay 1 + " at " + Time 1 + " on " + TicketClass</i>
Step 2. Fill template with values	"أريد حجز رحلة ذهاب من " + الرياض + " إلى " + الدوحة + " " + لطفلين و ثلاث بالغين + " " + يوم الخميس 2022/6/26 + " الساعة " + 9 مساء + " على " + الدرجة الأولى <i>"I want to book one-way ticket from " + Riyadh + " to " + Doha + " " + for two kids and three adults + " " + on Thursday 26/6/2022 + " at " + 9 PM + " on " + first class</i>
Step 3. Generate sample text	"أريد حجز رحلة ذهاب من الرياض إلى الدوحة لطفلين وثلاثة بالغين يوم الخميس 2022/6/26 الساعة 9 مساء على الدرجة الأولى" <i>I want to book one-way ticket from Riyadh to Doha for two kids and three adults on Thursday 22/6/2022 at 9 PM on first class</i>
Step 4. Represent sample as a sequence of tags	"000801029099333044077"

Table 5. Summary of the ABC-S dataset.

Domain	D1	D2	D3	D4	D5	D6	D7	Total
Name	Flight Booking	Hotel Booking	Cinema Booking	Football Match	Car Booking	Restaurant Booking	Clinic Booking	
Number of Used Entities	15	12	10	13	10	8	8	26
Number of Templates	12	5	5	7	6	7	10	52
Number of Samples	30,437	22,476	8,360	5,599	3,716	2,891	2,638	76,117

Table 6. The structure of the ABC-S dataset file.

Sentences	Tags	Domain	Domain-Template
مرحبا انا علي عوده الله بدي اسافر من البرازيل الى الصين بيوم 6 الشهر بتذكرة طيران من درجة اقتصادية <i>Hi, I'm Ali Odet Allah I want to travel from Brazil to China on the 6th on the economy class</i>	001010100010 233390077	1	D1T5
بدي انزل بفندق لو رويال في الإسكندرية واحجز 4 غرف نوم ل 3 اشخاص كبار <i>I want to stay in the LaRoyal hotel in Alexandria and book four rooms for three adults</i>	000161601017 1700999	2	D2T5
لو سمحت احجزلي تذاكر طيران ذهاب وعودة من الدمام الى القاهرة لشخصين بالغين و 3 اطفال على درجة اولي <i>Please book two-way tickets for me from Dammam to Cairo for two adults and 3 kids on the first class</i>	000908801029 9099077	1	D1T7

3.1.2 Collecting the ABC-C Dataset

To test the proposed model on real reservation samples, we proposed building the ABC-Collected

Dataset (ABC-C Dataset) which contains reservation samples collected from 38 volunteers. The volunteers were asked to freely write reservation samples in different domains without having prior knowledge about the generated synthetic samples in the ABC-S dataset. A total of 200 reservation samples were collected from the volunteers, with the distribution provided in Table 7. The tags for these sequences were generated manually. Table 8 displays some examples of these samples along with their corresponding tags.

Table 7. Distribution of reservation samples in the ABC-C dataset

Domain	D1	D2	D3	D4	D5	D6	D7
Number of Samples	30	24	29	29	27	30	31
Total	200						

Table 8. Structure of the ABC-C dataset file.

Sentences	Tags	Domain ID
مطلوب حجز موعد لعيادة الأسنان عند الدكتور نزار غنام الساعة الخامسة مساءً.	0 0 0 0 25 0 0 11 11 0 4 4	D1
مرحباً بدي حجز طاولة في مطعم جبري ع التراس لشخصين بدون ارجيلة.	0 0 0 0 0 0 18 0 0 9 0 0	D6
لو سمحت اريد حجز سيارة ميتسوبيشي باجبرو لتلاتة أيام وانكم تحضروها لموقعي في جبل الحسين شارع المجلد عمارة 30 ، و يتم تسليمها من قبلي في المكتب عنديكم.	0 0 0 0 0 21 21 23 23 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	D5

3.2 The Proposed Models

As we mentioned earlier, the main objective of this work is to develop and evaluate a named entity recognition model for Arabic booking chatbots that can be used with multiple domains. For this purpose, we propose two models that customize the AraBERT transformer as discussed in the following sub-sections.

3.2.1 The NERB Model

The first proposed model is referred to as the Named Entity Recognition Model for Booking Queries (NERB). The architecture of this model is shown in Figure 1. At the core of this model lies the AraBERTv0.2-base model. We chose to use this version of AraBERT, because it is trained on a larger dataset and has demonstrated better performance in the NER task [14]. We customized the AraBERT model by adding two new layers: a dropout layer with a rate of 0.3 to prevent overfitting and a linear output layer with 26 outputs corresponding to the number of entity categories in the ABC-S dataset.

The input to the AraBERT transformer is represented using three torch tensors: Input IDs, Input Mask and Segment IDs. Input IDs are tokens mapped into IDs. The Input Mask is a sequence of 1s and 0s, where 1s correspond to real tokens and 0s represent padding tokens. This mask is used to prevent the model from considering the padded tokens. Segment IDs utilize 1s and 0s to distinguish between two sentences. While BERT expects sentence pairs, in our case, the input query belongs to a single sentence, so the segment ID is essentially a sequence of 0s. AraBERT transforms these three torch tensors into a torch tensor with the dimensions of (Max Sequence Length, Token Vector Size = 60 and 768). The Max Sequence Length is set to 60 in NERB, while it is typically 512 in BERT. The value 768 corresponds to the hidden size dimensions of the encoder layer in the AraBERT model.

The output from the AraBERT transformer is subsequently used as input for the following layers, which include a dropout layer and a linear layer with 768 inputs and 26 outputs. The final output from this model is a torch tensor with dimensions of (Max Sequence Length, Number of Classes = 60 and 26). Following this, the output dimensions of the NERB model are reduced by selecting the classes with the highest probability and removing padding from the sequence, so that its length is equal to the length of the input vector.

The compilation of the proposed model was carried out using the same configuration as the base model. We employed the Adam optimizer with a learning-rate scheduler and utilized the cross-entropy

loss function. The model was trained using the ABC-S Dataset and we evaluated its performance using the accuracy score.

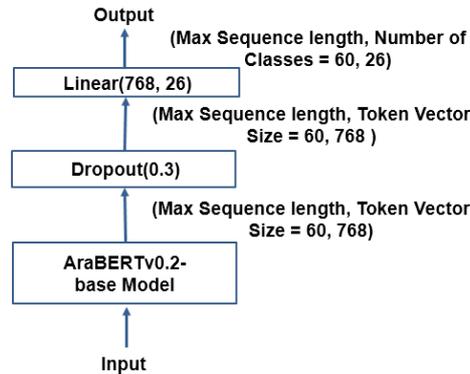


Figure 1. The architecture of the proposed NERB model.

For training, we used a maximum of 100 epochs for the training process. In each epoch, the model trains using the training set with a batch size of 32 and then predicts the tags of the validation set with a batch size of 8. We monitored the validation loss value and implemented early stopping if the validation loss did not improve for five consecutive epochs. The model with the lowest validation loss was adopted and utilized for testing. The model was trained on a GPU for 14 epochs, taking approximately 10 minutes per epoch and was then stopped due to the application of early stopping with a patience value set to 5. The ninth epoch recorded the lowest validation loss of $1.31e^{-06}$.

3.2.2 The DA-NERB Model

The second model that we propose is the Domain-aware Named Entity Recognition for Booking Queries (DA-NERB). The basic idea of this model is to extend the NERB model such that some prior knowledge about the domain is incorporated in the input to the model to investigate the effect of such knowledge on the overall performance. Figure 2 shows the architecture of the DA-NERB model. Technically, the DA-NERB model has an additional input which is the domain ID; a value between 1 and 7 as shown in Table 5. This input is fed into an embedding layer with 8 cells that transforms the single-value domain ID into an 8-element. The output of the embedding layer is then reshaped to match the size of the dropout layer and then fed into a concatenation layer that merges the embedding-layer output with the dropout layer to obtain a tensor of size (Max Sequence Length, Token Vector Size = 60 and 776). A linear output layer with 776 inputs and 26 outputs receives the concatenated tensor and outputs a tensor with size (Max Sequence Length, Number of Classes = 60 and 26). Also, the final-output dimensions will be reduced as we mentioned previously in the NERB model.

Compiling and training this model were performed using the same configuration of the NERB model using the ABC-S dataset. This model was trained using GPU for 12 epochs with approximately 10 minutes per epoch and then stopped due to the use of early stopping with the patience value set to 5. The seventh epoch has the lowest validation loss of 8.44×10^{-07} .

Since DA-NERB is aware of the domain to which the input query belongs, we have introduced a simple post-processing step to enhance its output predictions. The output of DA-NERB consists of a sequence of predicted tags for each word in the input. Essentially, if the domain is known, the predicted tags should contain values for entities that belong to that specific domain, alongside tags that correspond to 'Others'. Figure 3 illustrates the seven reservation domains in the ABC-S dataset, with entities that must be present (highlighted in red) in any input query within these domains, as well as entities that may be present (highlighted in black). Based on this information, the post-processing step involves replacing incorrectly predicted tags with '0', which is the tag for the 'Others' entity. For instance, if the input query belongs to the 'flight' domain, a tag corresponding to a 'Restaurant Name' entity should not be present, so it is replaced with '0'. This post-processing step can also be employed to identify any missing entities in the input. This functionality can be valuable when the model is deployed to interactively gather these entities from the user.

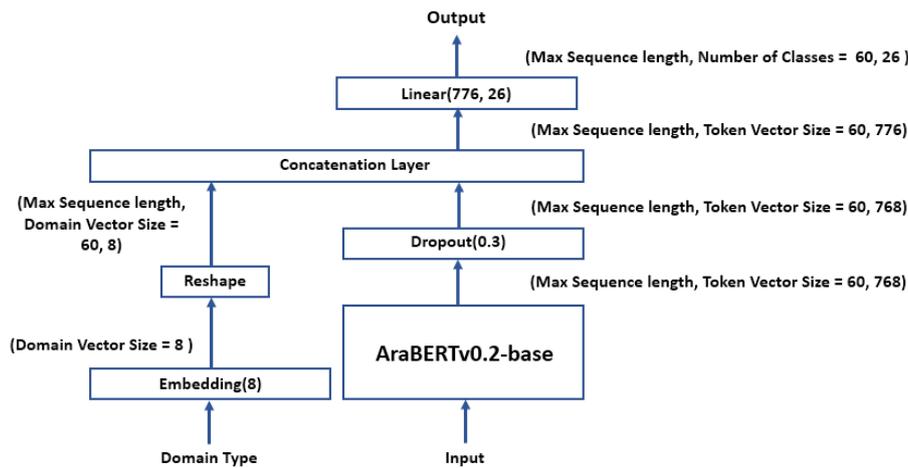


Figure 2. The architecture of the proposed DA-NERB model.

Classes \ Domain Type	Classes																									
	Other	City_1	City_2	Date_And_Day_1	Time_1	Date_And_Day_2	Time_2	Ticket_Class	Round_Trip	Numbers_Of_Tickets	Names	Doctor_Names	Phone	Team_1	Team_2	League	Hotel	Room	Restaurant	Movie	Cinema	Car_Name	Car_Model	Period_Time	Stadium	Clinic
Domain Type	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Airline Ticket Booking	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Hotel Booking	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Cinema Booking	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Football Match Booking	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Car Booking	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Restaurant Booking	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Clinic Appointment Booking	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Figure 3. The expected and required entity tags based on domain type.

4. RESULTS AND DISCUSSION

This section discusses the experimental platform used and presents the results obtained from experiments conducted on both the ABC-S and ABC-C datasets across various scenarios. The outcomes of these experiments encompass the performance of both the NERB and DA-NERB models. Furthermore, we provide a comparative analysis of the DA-NERB model's performance *versus* that of the NERB model when trained on a single reservation domain. We also compare our work with some previous studies. Finally, we outline the limitations of this work.

4.1 Experimental Setup

In our experiments, we utilized Kaggle notebooks to build, test the models, as well as to synthesize the ABC-S Dataset. The Kaggle platform offers a cost-free, browser-based environment that enables users to discover and share datasets, write, save and execute codes on Jupyter notebooks while utilizing powerful computing resources, including CPUs, GPUs and TPUs [25]. For our implementation and testing, we employed Python 3.7.10, Transformers 4.5.1 and PyTorch 1.9.1 libraries. The test set from the ABC-S along with the ABC-C datasets, was used to assess the proposed models. We evaluated the overall performance of the models using the accuracy metric.

4.2 The NERB Model Results

When evaluating the NERB model using the ABC-S dataset (test set), a classification accuracy score of 100% is achieved. This high accuracy is anticipated, because the AraBERT model is a transformer designed for the Arabic language and has undergone extensive training on a large corpus. However, the perfect accuracy score may indicate a notable similarity between the test set and the training set. To address this concern, we conducted an evaluation of the model using the ABC-C dataset, which we collected from 38 volunteers to ensure high diversity. This dataset comprises 200 booking tickets distributed across the seven domains.

Table 9 displays the classification report for evaluating the NERB model on the ABC-C dataset. This report includes precision, recall and F1-score metrics for the 25 entity categories. It's important to note that the "Phone" entity with a tag of 12 is not present, as none of the volunteers provided it in the collected reservation queries within the ABC-C dataset. The last column shows the support or the frequency of each entity in the dataset. In this table, we can see that three entities that have an F1-score value of 1.000, while many entities have a high value of F1-score regardless of their support in the dataset. For example, the "Stadium Name" entity has a support of 11 with an F1-score of 1.000 and the "City 1" entity has a support of 37 with an F1-score of 0.829. Overall, the NERB model achieved a classification accuracy of 95.5% on the ABC-C dataset.

Table 9. The classification report of the NERB model for the ABC-C dataset.

Label	Entity Name	Precision	Recall	F1-Score	Support
0	Others	0.985	0.964	0.974	1,798
3	Date and Day 1	0.984	0.966	0.975	262
9	Number of Tickets	0.927	0.953	0.940	214
4	Time 1	0.894	0.988	0.939	163
7	Ticket Class	0.916	0.950	0.933	80
23	Period	0.927	0.900	0.913	70
1	City 1	0.756	0.919	0.829	37
19	Movie	0.941	0.889	0.914	36
21	Car Name	1.000	0.853	0.921	34
16	Hotel	0.952	0.769	0.851	26
17	Room	0.677	0.920	0.780	25
25	Clinic Name	0.846	0.880	0.863	25
11	Doctor Name	0.950	1.000	0.974	19
5	Date and Day 2	0.864	1.000	0.927	19
13	First Team	1.000	1.000	1.000	17
14	Second Team	0.944	1.000	0.971	17
2	City 2	0.923	0.800	0.857	15
18	Restaurant	0.684	0.929	0.788	14
10	Name	0.619	1.000	0.765	13
20	Cinema	0.818	0.692	0.750	13
6	Time 2	1.000	1.000	1.000	13
24	Stadium Name	1.000	1.000	1.000	11
8	Round Trip	1.000	0.750	0.857	8
15	League	1.000	0.667	0.800	6
22	Car Model	0.333	1.000	0.500	1
Macro-average		0.878	0.912	0.881	Total: 2,936
Accuracy					95.5%

To better understand where the model fails in classification, Figure 4 shows the confusion matrix for the NERB model when evaluated using the ABC-C dataset.

Label	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
1 0 Others	1734	5	6	1	0	0	0	0	1	0	2	2	1	2	0	2	5	0	4	2	13	0	0	7	0	11
2 1 City1	3	34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3 10 Name	0	0	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4 11 Doctor Name	0	0	0	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5 13 First Team	0	0	0	0	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6 14 Second Team	0	0	0	0	0	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7 15 League	2	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8 16 Hotel	0	1	0	0	0	0	0	0	20	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9 17 Room	1	0	0	0	0	0	0	0	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
10 18 Restaurant	1	0	0	0	0	0	0	0	0	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11 19 Movie	3	0	1	0	0	0	0	0	0	0	32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12 2 City2	0	2	0	0	0	1	0	0	0	0	0	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13 20 Cinema	0	2	0	0	0	0	0	0	0	2	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0
14 21 Car Name	3	0	0	0	0	0	0	0	2	0	0	0	0	29	0	0	0	0	0	0	0	0	0	0	0	0
15 22 Car Model	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
16 23 Period of Time	1	0	0	0	0	0	0	0	2	0	0	0	0	0	0	63	0	0	2	0	0	0	0	0	0	2
17 24 Stadium Name	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0
18 25 Clinic Name	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	22	0	0	0	0	0	0	0	1
19 3 Date And Day1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	253	5	3	0	0	0	0	0
20 4 Time1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	151	0	0	0	0	0	0	0
21 5 Date and Day2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	19	0	0	0	0	0
22 6 Time2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	0	0	0	0	0
23 7 Ticket Class	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	76	0	0
24 8 RoundTrip	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	1
25 9 Number Of Tickets	6	1	1	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	204

Figure 4. Confusion matrix for the NERB-model predictions on the ABC-C dataset.

From this matrix, we can see that a total of 64 instances of the "Others" entity were classified as one

of the defined entities in the dataset. Also, the low precision of the "Name" entity, which has a support of 13, is due to the fact that 6 of the "Others" entities are classified as "Name" entities; despite the fact that the "Name" entity has a recall of 1. On the other hand, the "League" entity with a support of 6 has a recall value of 0.667 and a precision value of 1.000.

4.3 The DA-NERB Model Results

Like the evaluation results of the NERB model on the ABC-S dataset, the DA-NERB model achieved a classification accuracy score of 100%. Therefore, we turned to using the ABC-C dataset for evaluation. In Table 10, you can observe the classification report for evaluating the DA-NERB model on the ABC-C dataset, where it achieved a classification accuracy of 96.4%, surpassing the NERB model's performance. Furthermore, the table highlights that the DA-NERB model has improved F1-score values for many entities in comparison to the NERB model. For example, the 'Name' entity, which had an F1-score of 0.765 for the NERB model, reached a perfect F1-score of 1.000 with the DA-NERB model. Nevertheless, this model obtained lower F1-scores for a few entities, such as the 'First Team' and 'Cinema' entities.

Table 10. The classification report of the DA-NERB model for the ABC-C dataset.

Label	Entity Name	Precision	Recall	F1-Score	Support
0	Others	0.983	0.973	0.978	1,798
3	Date and Day 1	0.960	1.000	0.979	262
9	Number of Tickets	0.928	0.963	0.945	214
4	Time 1	0.988	0.969	0.978	163
7	Ticket Class	0.948	0.913	0.930	80
23	Period	0.986	0.971	0.978	70
1	City 1	0.897	0.946	0.921	37
19	Movie	0.897	0.972	0.933	36
21	Car Name	1.000	0.647	0.786	34
16	Hotel	1.000	0.808	0.894	26
17	Room	0.759	0.880	0.815	25
25	Clinic Name	0.880	0.880	0.880	25
11	Doctor Name	0.905	1.000	0.950	19
5	Date and Day 2	0.950	1.000	0.974	19
13	First Team	0.944	1.000	0.971	17
14	Second Team	1.000	1.000	1.000	17
2	City 2	0.929	0.867	0.897	15
18	Restaurant	0.778	1.000	0.875	14
10	Name	1.000	1.000	1.000	13
20	Cinema	0.526	0.769	0.625	13
6	Time 2	1.000	0.923	0.960	13
24	Stadium Name	0.917	1.000	0.957	11
8	Round Trip	0.857	0.750	0.800	8
15	League	0.714	0.833	0.769	6
22	Car Model	0.500	1.000	0.667	1
Macro-average		0.890	0.923	0.899	Total: 2,936
Accuracy					96.4%

Figure 5 displays the confusion matrices for the ABC-C dataset of the DA-NERB model. According to this figure, the "Car Name" entity has a precision value of 1.000 while this entity has a low recall value of 0.647, because the model classifies the "Car Name" entity as the other entities. The DA-NERB model has notably enhanced precision values for many entities, as seen in the case of the 'City 1' entity, with a precision value of 0.756 for the NERB model and an improved value of 0.897 for the DA-NERB model. Additionally, the DA-NERB model has improved recall values for various entities, such as the 'Movie' entity, which has a recall value of 0.889 for the NERB model and an improved value of 0.972 for the DA-NERB model. However, there were a few instances where the results were opposite, as observed in the case of the 'Car Name' entity, where the NERB model achieved a recall value of 0.853, while the DA-NERB model had a lower recall value of 0.647. Overall, the DA-NERB model demonstrated an improvement in the F1-score for numerous entities.

As we mentioned in Sub-section 3.2.2, the DA-NERB model is aware of the reservation domain of the input query; hence, we can apply the post-processing step that we discussed earlier to refine the classified tags and identify the missing tags in the query based on the domain. Effectively, applying the post-processing step improved the DA-NERB model and increased the accuracy to 96.9% when the ABC-C dataset was considered.

Label	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1 0 Others	1750	2	0	0	0	0	2	0	1	2	2	1	9	0	0	1	0	3	4	2	0	0	4	1	14
2 1 City1	1	35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
3 10 Name	0	0	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4 11 Doctor Name	0	0	0	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5 13 First Team	0	0	0	0	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6 14 Second Team	0	0	0	0	0	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7 15 League	0	0	0	0	1	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8 16 Hotel	0	0	0	0	0	0	21	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9 17 Room	3	0	0	0	0	0	0	0	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10 18 Restaurant	0	0	0	0	0	0	0	0	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11 19 Movie	1	0	0	0	0	0	0	0	0	35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12 2 City2	0	2	0	0	0	0	0	0	0	0	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13 20 Cinema	0	0	0	0	0	0	0	2	1	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14 21 Car Name	11	0	0	0	0	0	0	0	0	0	0	0	22	1	0	0	0	0	0	0	0	0	0	0	0
15 22 Car Model	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
16 23 Period of Time	0	0	0	0	0	0	0	0	0	0	0	0	0	0	68	0	0	2	0	0	0	0	0	0	0
17 24 Stadium Name	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0
18 25 Clinic Name	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	22	0	0	0	0	0	0	0	1
19 3 Date And Day1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	262	0	0	0	0	0	0	0
20 4 Time1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	4	158	0	0	0	0	0	0
21 5 Date and Day2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	19	0	0	0	0	0
22 6 Time2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	12	0	0	0	0
23 7 Ticket Class	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	73	0	0	0
24 8 RoundTrip	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	1	0
25 9 Number Of Tickets	6	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	208	0

Figure 5. Confusion matrix for the DA-NERB model predictions on the ABC-C dataset.

Table 11. Sample from the ABC-C dataset and the outputs of the NERB, DA-NERB and the post-processing step.

Words in Query	True Label		NERB Output		DA-NERB Output		Post-processing	
	Tag	Entity	Tag	Entity	Tag	Entity	Tag	Entity
مسا	0	Other	<u>4</u>	Time 1	0	Other	0	Other
الخير،	0	Other	0	Other	0	Other	0	Other
بأدر	0	Other	0	Other	0	Other	0	Other
احجز	0	Other	0	Other	0	Other	0	Other
الجي	21	Car Name	<u>17</u>	Room	<u>0</u>	Other	0	Other
كلاس	21	Car Name	<u>17</u>	Room	<u>0</u>	Other	0	Other
لثلاث	23	Period	23	Period	23	Period	23	Period
أيام	23	Period	23	Period	23	Period	23	Period
بس	0	Other	0	Other	0	Other	0	Other
ضروري	0	Other	0	Other	0	Other	0	Other
كثير	0	Other	0	Other	0	Other	0	Other
تكون	0	Other	0	Other	0	Other	0	Other
لبلاك	0	Other	<u>10</u>	Customer Name	<u>9</u>	Number of Tickets	<u>0</u>	Other
لأنها	0	Other	0	Other	0	Other	0	Other
طلب	0	Other	0	Other	0	Other	0	Other
المدام،	0	Other	<u>10</u>	Customer Name	0	Other	0	Other
ممنونكم	0	Other	0	Other	0	Other	0	Other

To illustrate the concept of the post-processing step, Table 11 provides the output of the NERB, DA-NERB and the post-processing step for a sample from the ABC-C dataset. The sample is “مسا الخير، بأدر ” (Good Evening, I want to book the G class for three days, but it has to be black, because my wife asked for it, thank you.), which is a query for a car rental. The NERB output has five errors, while the DA-NERB has three errors, which are shown with underlines in the table. Both models failed to correctly tag the car model “الجي كلاس” (G Class) and the car color “لبلاك” (black). However, applying the post-processing step to the DA-NERB output corrected the classification of the car color to 0, which matches the true label of “Others”. Furthermore, the post-processing step identified missing tags in the query, such as tags 3 (Start Date), 4 (Start Time), 5 (End Date), 6 (End Time), 21 (Car Name) and 22 (Car Model). These missing tags are expected to be acquired by prompting the user interactively once the model is deployed.

4.4 Comparison with the Single-Domain

For a more comprehensive and fair comparison, we conducted an experiment to evaluate the performance of the DA-NERB model when compared to the performance of the NERB when it is trained on a single domain. For this purpose, we trained seven versions of the NERB model, each tailored to one of the seven domains. Technically, each of these single-domain models was trained using a sub-set of the samples from the ABC-S dataset specific to the domain of that model. For testing, we assessed these individual models alongside the DA-NERB model using the domain-specific samples from the ABC-C dataset. Table 12 provides a comparison between the single-domain models and the DA-NERB model with post-processing. It is evident that the accuracy of the DA-NERB model consistently surpasses that of each of the single-domain models.

Table 12. Comparison between the DA-NERB model with post-processing and the single-domain models.

Domain Type	ABC-C Dataset Samples/Entities	Single-domain NERB Models (Without Post-processing)	DA-NERB Model (With Post-processing)
		Accuracy	Accuracy
Flight Booking	30 / 494	96%	97%
Hotel Booking	24 / 393	89%	96%
Cinema Booking	29 / 391	95%	96%
Football Match Booking	29 / 375	94%	98%
Car Booking	29 / 401	93%	95%
Restaurant Booking	30 / 446	94%	98%
Clinic Appointment Booking	31 / 436	95%	98%

4.5 Comparison with Previous Works

Fadhil [12] focused on the patient follow-up domain, Al-Ajmi and Al-Twairesh [10] handled the flight booking domain and most of the Arabic NER works focused on a single domain, whereas this work addressed seven different domains.

Shaker *et al.* [22] introduced an Arabic NER dataset encompassing 9 categories of named entities, comprising over 36,000 records, with texts spanning 7 different domains. This work presented the ABC-S Dataset, which comprises 76,117 reservation samples spanning 7 different domains and encompassing 26 categories of named entities.

We fine-tuned AraBERTv0.2 on the ABC-S dataset that supports 26 different entity classes; in contrast, Antoun *et al.* [18] fine-tuned AraBERTv0.1 on ANERcorp which contains entities belonging only to 4 different classes.

4.6 Limitations

One of the constraints in this work is the utilization of the DA-NER model to recognize novel entities that fall outside the reservation domains on which it was previously trained. Another constraint in this work is the dialect used, especially when the reservation request is entered in a dialect other than the Levantine dialect. In such cases, the model may be able to recognize certain entities, such as time and date, but it may face difficulty in recognizing other entities that rely on the text context.

5. CONCLUSION

In this paper, we proposed the design and evaluation of a named entity recognition model for Arabic chatbots. We achieved this by fine-tuning and extending the state-of-the-art AraBERT to recognize booking entities and information from unstructured Arabic reservation queries. Due to the scarcity of datasets for reservation queries in Arabic, we proposed the synthesis of the ABC-S dataset and the collection of the ABC-C datasets for training and testing the proposed models. These datasets contain reservation queries for seven domains.

Effectively, we proposed two models: the NERB and DA-NERB. The NERB model is basically the AraBERT model that is modified by adding an output layer of 26 outputs and trained on the ABC-S

dataset. On the other hand, the DA-NERB extends the NERB model by including an additional input for the domain type and an embedding layer to incorporate prior knowledge about the domain.

Experimental evaluation of these models proved the ability of the NERB and DA-NERB models to recognize the entities in the queries in the ABC-C dataset with an accuracy of 95.5% and 96.4%, respectively. Furthermore, we proposed improving the DA-NERB classification accuracy using a post-processing step utilizing the fact that the domain type is known. The classification accuracy of the DA-NERB increased to 96.9% using this post-processing.

One major contribution of this work is building a specialized dataset for Arabic unstructured reservation texts. Specifically, we built the Arabic Booking Chatbot Dataset (ABC-S Dataset), which contains 76,117 reservation samples encompassing 26 entity types for seven different reservation domains. Also, we collected samples for booking tickets from 38 volunteers. These datasets are publicly available to other researchers to contribute to the development of Arabic chatbots.

In the future, we may eliminate the need for post-processing by utilizing other pre-trained language models, such as MARBERT [27] and AraT5 [26]. Additionally, it is possible to train the model on additional dialects. Moreover, we are looking to integrate our model with an available library to build a chatbot reservation system and then enable the users to book tickets in Arabic text and set some conditions and restrictions for booking tickets, in order to avoid misunderstanding when using the booking chatbot.

REFERENCES

- [1] E. H. Almansor and F. K. Hussain, "Survey on Intelligent Chatbots: State-of-the-art and Future Research Directions," Proc. of Conf. on Complex, Intelligent and Software Intensive Systems (CISIS 2019), vol. 993, pp. 534–543, 2020.
- [2] M. Al-Ayyoub et al., "Deep Learning for Arabic NLP: A Survey," JOCSA, vol. 26, pp. 522-531, 2018.
- [3] S. AlHumoud, A. Al Wazrah and W. Aldamegh, "Arabic Chatbots: A Survey," IJSCSA, vol. 9, no. 8, pp. 535-541, 2018.
- [4] M. Mnasri, "Recent Advances in Conversational NLP: Towards the Standardization of Chatbot Building," DOI: 10.48550/arXiv.1903.09025, Clermont-Ferrand, France, 2019.
- [5] Infobip, "The Intelligent Chatbot Building Platform," [Online], Available: <https://www.infobip.com>, 2021.
- [6] T. Alshareef and M. A. Siddiqui, "A seq2seq Neural Network Based Conversational Agent for Gulf Arabic Dialect," Proc. of the 2020 21st IEEE Int. Arab Conf. on Information Technology (ACIT), pp. 1-7, Giza, Egypt, 2021.
- [7] Y. Saoudi and M. M. Gammoudi, "Trends and Challenges of Arabic Chatbots: Literature Review," Jordanian Journal of Computers and Information Technology (JJCIT), vol. 9, no. 3, pp. 261-286, 2023.
- [8] Class Central, "Microsoft Bot Framework and Conversation as a Platform," [Online], Available: <https://www.classcentral.com/course/edx-microsoft-bot-framework-and-conversation-as-a-platform-11325>, 2021.
- [9] Chatbots Life, "Best Chatbot Development Frameworks | RASA | IBM Watson | Dialogflow," [Online], Available: <https://chatbotslife.com/best-chatbot-development-frameworks-rasa-ibm-watson-dialogflow-e2792f9363eb>, 2019.
- [10] A. H. Al-Ajmi and N. Al-Twairesh, "Building an Arabic Flight Booking Dialogue System Using a Hybrid Rule-based and Data Driven Approach," IEEE Access., vol. 9, pp. 7043-7053, Jan. 2021.
- [11] D. Al-Ghadhban and N. Al-Twairesh, "Nabiha: An Arabic Dialect Chatbot," IJACSA, vol. 11, no. 3, pp. 452-459, 2020.
- [12] A. Fadhil, "Ollobot-Towards a Text-based Arabic Health Conversational Agent: Evaluation and Results," Proc. of the Int. Conf. on Recent Advances in Natural Language Processing (RANLP 2019), pp. 295-303, Varna, Bulgaria, 2019.
- [13] G. Mesnil et al., "Using Recurrent Neural Networks for Slot Filling in Spoken Language Understanding," IEEE/ACM Transactions on Audio, Speech and Language Processing, vol. 23, no. 3, pp. 530-539, DOI:10.1109/TASLP.2014.2383614, Mar. 2015.
- [14] W. Antoun, F. Baly and H. Hajj, "AraBERT: Transformer-based Model for Arabic Language Understanding," Proc. of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection (LREC2020), pp. 9-15, Marseille, France, 2020.
- [15] J. Devlin, M. W. Chang, K. Lee and K. Toutanova, "Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding," Proceedings of NAACL-HLT2019, pp. 4171–4186, Minneapolis, Minnesota, June2-June7, 2019.
- [16] R. Horev, "BERT Explained: State of the art Language Model for NLP," Towards Data Science,

- [Online], Available: <https://towardsdatascience.com/bert-explained-state-of-the-art-language-model-for-nlp-f8b21a9b6270>, 2018.
- [17] D. Shulga, "BERT to the Rescue," Towards Data Science, [Online], Available: <https://towardsdatascience.com/bert-to-the-rescue-17671379687f>, 2019.
- [18] W. Antoun, "Aub-mind/arabert: Pre-trained Transformers for the Arabic Language Understanding and Generation (Arabic Bert, Arabic GPT2, Arabic Electra)," GitHub, Edited by M. Al Salti et al. AUB MIND, Beirut, Lebanon, [Online], Available: <https://github.com/aub-mind/arabert>, 2020.
- [19] A. Youssef, M. Elattar and S. R. El-Beltagy, "A Multi-embeddings Approach Coupled with Deep Learning for Arabic Named Entity Recognition," Proc. of the 2020 2nd IEEE Novel Intelligent and Leading Emerging Sciences Conf. (NILES), pp. 456-460, Giza, Egypt, 2020.
- [20] E. Taher, S. A.Hoseini and M.Shamsfard, "Beheshti-NER: Persian Named Entity Recognition Using BERT," Proc. of the 1st Int. Workshop on NLP Solutions for Under Resourced Languages (NSURL 2019) Co-located with ICNLSP 2019, pp. 37-42, Trento, Italy, 2020.
- [21] B. A. Benali, S. Mihi, N. Laachfoubi and A. A. Mlouk, "Arabic Named Entity Recognition in Arabic Tweets Using BERT-based Models," Procedia Computer Science, vol. 203, pp. 733-738, 2022.
- [22] A. Shaker, A. Aldarf and I. Bessmerntny, "Using LSTM and GRU with a New Dataset for Named Entity Recognition in the Arabic Language," arXiv: 2304.03399, DOI: 10.48550/arXiv.2304.03399, 2023.
- [23] X. Qu, Y. Gu, Q. Xia et al., "A Survey on Arabic Named Entity Recognition: Past, Recent Advances and Future Trends," arXiv: 2302.03512, DOI: 10.48550/arXiv.2302.03512, 2023.
- [24] B. Saddar and R. Saddar, "Boshra-sadder/Arabic Booking Chatbot," GitHub, [Online], Available: <https://github.com/Boshra-sadder/Arabic-Booking-Chatbot>, Amman, Jordan, 2021.
- [25] G.Yufeng, "Introduction to Kaggle Kernels," Towards Data Science, [Online], Available: <https://towardsdatascience.com/introduction-to-kaggle-kernels-2ad754ebf77>, 2017.
- [26] E. M. B. Nagoudi, A. Elmadany and M. Abdul-Mageed, "TURJUMAN: A Public Toolkit for Neural Arabic Machine Translation," arXiv: 2206.03933, DOI: 10.48550/arXiv.2206.03933, 2022.
- [27] B. AIKhamissi et al., "Adapting MARBERT for Improved Arabic Dialect Identification: Submission to the NADI," arXiv: 2103.01065, DOI: 10.48550/arXiv.2103.01065, 2021.

ملخص البحث:

لقد أصبحت أنظمة التّحاور أساسية في الأونة الأخيرة في العديد من المجالات، بدءاً من خدمة الزّبائن والحصول على المعلومات وانتهاءً بالترفيه، ومما يجدر ذكره أنّ استخدام أنظمة التّحاور من شأنه أن يخفّض الكلفة التشغيلية ويقلّل الأخطاء البشرية بالإضافة الى توفيره للخدمة في جميع الأوقات.

يقدّم هذا البحث نموذجاً لتمييز الكينونات المسماة (NER) لتصميم وتطبيق وتقييم نظامٍ للتّحاور باللّغة العربية يتعلّق بتقديم خدمات الحجز، ويُركّز على حجز التذاكر والمواعيد في سبعة مجالات مختلفة، وهي الرحلات الجوية، والفنادق، ودور السينما، ومباريات كرة القدم، والسيارات، والمطاعم، والعيادات. ويمهّد هذا البحث الطّريق لتطوير المزيد من أنظمة التّحاور باللّغة العربية التي بإمكانها أن تدعم العديد من مجالات الحجز. وتستخدم هذه الورقة تعلّم الآلة العميق وتعلّم النقل لإنجاز هذه المهمة. وعلى وجه التّحديد، يستخدم النّمودج المقترح في هذه الدّراسة النّمودج الاساسي AraBERTv0.2 لتدريب وتقييم أداء النّمودج. اعتمدت الدّراسة على إنشاء مجموعة بيانات باللّغة العربية تشمل (76117) عيّنة حجز، وتغطي سبعة مجالات مختلفة، وتتضمّن (26) صنفاً من الكينونات المسماة. بالإضافة الى ذلك، تمّ جمع مجموعة بياناتٍ من متطوّعين لتقييم النّمودج المقترح بعد تدريبه. ويُشار إلى أنّ مجموعتي البيانات المذكورتين مكتوبتان باللّغة العربية المتداولة في المشرق. وقد بلغت دقّة النّمودج المقترح 100% و 96.9% عند تطبيقه على مجموعتي البيانات المشار إليهما أعلاه. والمتوقّع أن يدعم النّمودج المقترح البحث العلمي المتعلّق بأنظمة التّحاور باللّغة العربية.

RADIO FREQUENCY IDENTIFICATION NETWORK PLANNING USING ASCENDANT HIERARCHICAL CLUSTERING ALGORITHM

Lamri Sayad

(Received: 27-Sep.-2023, Revised: 2-Dec.-2023, Accepted: 23-Dec.-2023)

ABSTRACT

Radio Frequency Identification (RFID) is a suitable technology for several application scenarios. However, it faces many issues that need to be addressed. RFID Network Planning (RNP) is one of these challenges, which is studied in this paper. It consists in defining the optimal placement of RFID readers and their communication range. In this study, the RNP issue is considered as a clustering problem. Therefore, it is solved using Ascendant Hierarchical Clustering (AHC) algorithm, such that each cluster head represents a reader. The advantage of our proposed approach is that, in addition to optimizing the RNP, it is capable to determine the optimal number of readers needed to get optimal performances. Simulation results demonstrate that the proposed AHC algorithm performs better than two efficient RNP algorithms from the literature: Particle Swarm Optimization (PSO) and Leaders and Followers (LaF) algorithms. Hence, it is observed that the AHC algorithm is more efficient, scalable and runs faster than PSO and LaF.

KEYWORDS

RFID, Network planning, Hierarchical clustering, Agglomerative hierarchical clustering, Reader placement.

1. INTRODUCTION

Radio Frequency Identification (RFID) is a widely used wireless technology that involves the use of radio waves to track and identify items, products and individuals. It is considered as one of the key technologies in 21st century [1]. Hence, according to [2], the RFID market is expected to reach more than 36 billion US Dollars by 2030. A typical RFID system is composed of three kinds of components: tags, readers and a host computer. RFID tags are small electronic devices consisting on a chip and an antenna. Each tag is embedded in or attached to the object to track and stores data that can be read by an RFID reader. Based on their functionality, RFID tags can be of three types: passive, active and semi-passive. Passive RFID tags are not equipped with a power source; hence, they are powered and activated by the energy transmitted by the RFID reader. On the other hand, active RFID tags, unlike passive tags, do not rely on the energy transmitted by the RFID reader, but they have their own power source. Semi-passive RFID tags still depend on the energy sent by the reader. However, this energy is used only for communication and they use their own power source to power their internal circuitry. An RFID reader is a device used to collect information from the tags within its communication range. The role of the host computer, as a component of an RFID system, is to process, analyze and store the data received from the RFID readers.

The fact that RFID systems eliminate the need for direct physical contact to track and identify objects makes them very useful in a wide range of application fields, including:

- Supply Chain Management: it is used in inventory management in warehouses and distribution centres to improve inventory accuracy.
- Transportation and Logistics: RFID technology is also involved in many transportation and logistics operations, such as tracking and tracing goods, managing shipping containers and monitoring vehicle movements.
- Retail: it is used for item-level tracking.
- Healthcare: RFID systems can also be used to enhance patient safety by monitoring patients and tracking medical equipment [3] [4].
- Manufacturing: it is used to improve the efficiency and enable better quality control in asset tracking, inventory management and process control in manufacturing facilities.

- Access Control and Security: RFID cards can provide a suitable tool to control and secure the access to buildings and restricted areas.

Furthermore, RFID technology can be integrated with an IoT system by connecting RFID systems to the internet.

Even though RFID technology is widely used in many application areas, several challenges and issues are still open and new solutions are needed and/or current solutions need improvements. RFID Network Planning (RNP) is one of these issues that need to be considered and addressed before the deployment of such networks. The RNP issue refers to the process of defining the appropriate network parameters, such that the performance of the network will be optimized. Usually, this planning problem involves three tasks: i) defining the optimal number of readers, ii) defining the adequate readers' positions and iii) defining the optimal transmission power of every reader. The main objectives when dealing with the RNP issue are maximizing tag coverage, minimizing interferences and minimizing cost (both capital and operational). It has been proven in the literature that the RNP issue is NP-hard [5].

In this paper, we apply the Ascendant Hierarchical Clustering (AHC) algorithm, also known as agglomerative hierarchical clustering, to solve this RNP issue. AHC algorithm is a flexible approach for dataset clustering and has applications in various fields, such as biology, social sciences and marketing, among others. Hence, this work aims to optimize five parameters: number of readers, tag coverage, interferences, power and power loss.

Several algorithms have been proposed to tackle the RNP issue. Most of them are based on metaheuristic algorithms. Guan et al. [5] have proposed the first algorithm that solves the RNP issue using genetic algorithm. Particle Swarm Optimization (PSO) was also applied in [1, 6, 7] to solve the RNP issue. In [7], a multi-swarm optimizer, called PS²O, has been proposed to solve the RNP issue by considering the following objectives: tag coverage, reader collision, economic efficiency and load balance. Hence, multiple populations of the PSO algorithm are combined in a hierarchical interaction topology. Furthermore, another PSO-based approach has been proposed in [6]. This algorithm, called HPSO-RNP, is a combination between PSO with K-means and virtual force algorithms. In [8], Genetic algorithm has been applied to solve the RFID network-planning problem based on directional and omnidirectional antenna readers. In this approach, a reader-elimination method has been introduced in order to remove unnecessary readers (redundant readers) from the network. Two metrics have been considered while addressing this issue: tag interference and reader-load balance. In addition, Zhao et al. [9] have applied a multi-objective firefly algorithm (FA) to solve the RNP issue with uncertainty. Similarly, authors have considered the optimization of three objectives: tag coverage, deployment cost and interference level. An improved version of FA algorithm has been used in [10] to address the same RNP issue. Moreover, another algorithm called Mayfly Sparrow Search in [11] tries to optimize four objectives : tag coverage, interference, power and load balance. M. Shokouhifar [12] has proposed a combination of whale optimization algorithm and simulated annealing to solve the RNP issue used to track medical assets in hospital environments. This solution considers the optimization of five parameters : coverage, cost, interference, collision and power. Recently, artificial neural networks [13], leaders and followers algorithm [14] and biogeography-based optimization [15] have been applied to solve the RNP issue.

Although the first works addressing the RNP issue have been designed since about two decades and several algorithms and solutions have been proposed to solve this issue, the need for new solutions is still required to address multiple objectives and the scalability of the RNP solutions. Hence, in this paper, the RNP issue is seen as a clustering problem, such that: the tags are the points to split into clusters, clusters are the RFID readers and the centres of clusters are the positions of the readers. Accordingly, we propose, in this paper, to solve the RNP issue using Ascendant Hierarchical Clustering (AHC) algorithm. The idea behind this algorithm is very simple. At the beginning, every tag is considered as a cluster. Then, at every iteration, the closest clusters are merged into larger clusters. Unlike existing algorithms, AHC deals with the RNP issue very fast and has a linear execution-time complexity.

The contributions of this paper are summarized as follows :

1. We propose a model for RFID system and a mathematical formulation to the multi-objective RNP (MORNP) problem. Hence, five optimization objectives have been considered: number of readers, tag coverage, interferences, power and power loss.

2. We design an approach, called AHC-MORNPN, based on the AHC algorithm to deal with the MORNPN problem.
3. We have implemented, evaluated and then compared the performance of the proposed approach with those of two RNP algorithms from the literature: PSO [7] and LaF [14].

The remainder of this paper is organized as follows: In the next section, the model of the system will be given, then a mathematical formulation of the RNP problem will be proposed. Section 3 will provide an explanation of the proposed Ascendant Hierarchical Clustering (AHC) algorithm and how it is applied to solve the considered RNP problem. The evaluation process and the simulation results are presented and described in Section 4. Finally, some conclusions and perspectives are provided in Section 5.

2. SYSTEM MODEL AND PROBLEM FORMULATION

This section will describe how the system is modelled. Then, it describes the mathematical formulation of the problem. Therefore, let us first describe the symbols used in this paper, as illustrated in Table 1.

Table 1. List of symbols and parameters used in this paper.

<i>Notation</i>	<i>Description</i>
<i>Inputs</i>	
t_i	The i -th tag
r_j	The j -th reader
n	Total number of tags
m	Maximum number of readers
Γ	The deployment area
W	Deployment area width
H	Deployment area height
B_t	The t -th boundary of the deployment area
(x_t, y_t)	Coordinates of the t -th tag
ρ_{Min}	Minimum communication radius of readers
ρ_{Max}	Maximum communication radius of readers
<i>Model variables</i>	
d_{rt}	Euclidean distance between tag t and reader r
δ_{rt}	Defines whether tag t is inside the communication range of r
φ_{rj}	Defines whether a reader r is inside the interference range of another reader j
C_{rt}	Defines whether tag t is covered by reader r
<i>Outputs</i>	
k	Number of selected readers
(x_r, y_r)	Coordinates of the r -th reader
σ_j	Interference radius of the j -th reader
ρ_j	Communication radius of the j -th reader
<i>Objective variables</i>	
$DR(G)$	Number of deployed readers in the RFID network represented by the graph G
$Cov(G)$	Coverage ratio of the RFID network represented by the graph G
$Interf(G)$	Interference of the RFID network represented by the graph G
$Pow(G)$	Total power consumed by the RFID network represented by the graph G
$PLoss(G)$	Power lost by the RFID network represented by the graph G
<i>AHC symbols</i>	
c_i	i -th cluster
S_{ij}	Similarity measure between clusters c_i and c_j .

The RFID system is designed under the following assumptions:

- All tags are passive due to their low cost [16], meaning that they are powered by the energy transported by the radio signals transmitted by the RFID readers.
- The RFID system is deployed on a 2D rectangular area.
- The RFID readers can be placed anywhere on the deployment area.
- Every node of the network (reader or tag) has one omnidirectional antenna.

- The RFID readers are heterogeneous in terms of communication range.

2.1 System Model

We consider a 2D rectangular deployment area Γ with a width W and a height H . The RFID network is represented as a graph $G(V, E)$, where V is the set of vertices or nodes and E is the set of edges. Additionally, there are two kinds of nodes in the network: tags and readers. Hence, $V = T \cup R$, such that:

- The set $T = \{t_1, t_2, \dots, t_n\}$ of n passive tags attached to the objects to be tracked or identified. We consider (x_t, y_t) as the coordinates of the t -th tag ($1 \leq t \leq n$).
- The set $R = \{r_1, r_2, \dots, r_m\}$ of available RFID readers. The aim of the RNP issue is to find the best number of readers k ($k \leq m$) and their positions (x_r, y_r) on the deployment area Γ ($1 \leq r \leq k$). In addition, it also aims to find the best interrogation range ρ_r of every reader r to meet the requirements, where $\rho_{Min} \leq \rho_r \leq \rho_{Max}$, $r \leq k$.

2.2 Problem Formulation

In this paper, the RNP problem considers the optimization of five objectives that are formulated as follows:

2.2.1 Coverage

Mainly, the most important requirement when addressing the RNP issue is to cover the maximum number of tags. Hence, we define the coverage $Cov(G)$ of the graph G , representing our RFID system, as follows:

$$Cov(G) = \frac{\sum_{t=1}^n (\sum_{r=1}^k (C_{rt}))}{n} \quad (1)$$

where:

$$C_{rt} = \begin{cases} 1 & \text{if tag } t \text{ is covered by reader } r. \\ 0 & \text{Otherwise} \end{cases} \quad (2)$$

and $\sum_{r=1}^k (C_{rt}) \leq 1, \forall t \in \{1, \dots, n\}$.

2.2.2 Interference

In addition to coverage, the aim of this work is also to minimize the interferences caused by the nodes of the RFID network G . Interferences are proportional to the density of the network [17]. Hence, the higher the density, the greater the number of interferences. In RFID, there are two kinds of interferences: interference between two or more close readers situated inside their communication range (reader interference) and interference caused by two or more readers covering the same tag (tag interference) [11]. Therefore, Equation (3) defines how interference metric is computed.

$$Interf(G) = \frac{\sum_{t=1}^n TI(t) + \sum_{r=1}^k RI(r)}{m + n} \quad (3)$$

where $TI(t)$ defines whether interferences occurred at the tag t . $TI(t)$ is computed according to the following equation:

$$TI(t) = \begin{cases} 1 & \text{if } (\sum_{r=1}^k \delta_{rt}) > 1, \\ 0 & \text{Otherwise} \end{cases} \quad (4)$$

$$\delta_{rt} = \begin{cases} 1 & \text{if } d_{rt} \leq \rho_r, \\ 0 & \text{Otherwise} \end{cases} \quad (5)$$

and $RI(r)$ defines whether interferences occurred at the reader r and is computed according to the following equation:

$$RI(r) = \begin{cases} 1 & \text{if } (\sum_{j=1, j \neq r}^k \varphi_{rj}) > 0, \\ 0 & \text{Otherwise} \end{cases} \quad (6)$$

$$\varphi_{rj} = \begin{cases} 1 & \text{if } d_{rj} \leq \sigma_j \\ 0 & \text{Otherwise} \end{cases} \quad (7)$$

2.2.3 Power

The transmitted power by all readers needs also to be minimized when solving the RNP problem. Based on the Friis transmission equation [18], the power received at a tag is calculated as follows:

$$P_{rec} = P_{tr} \cdot G_{rec} \cdot G_{tr} \cdot \left(\frac{\lambda}{4\pi R}\right)^2 \quad (8)$$

where P_{rec} is the received power, P_{tr} is the transmitted power, G_{rec} is the receiver-antenna gain, G_{tr} is the transmitter-antenna gain, λ is the wavelength and R is the distance between the reader and the tag.

According to eq. (8), the only factor that can affect the power received by a tag or a reader is the distance between them, which represents the radius of the reader's antenna. Therefore, to optimize the transmission power of a reader while dealing with the RNP issue, it is more convenient to minimize the radius of the reader ρ_j , where $\rho_{Min} \leq \rho_j \leq \rho_{Max}$ $j \leq k$.

Then, in this work, the total transmission power of an RFID network is defined as a function of the transmission radius, as shown in the following equation:

$$POW(G) = \frac{\sum_{r=1}^k \rho_r}{m * \rho_{Max}} \quad (9)$$

2.2.4 Power Loss

There is another way to save additional power in addition to the power saved previously. In this work, we also consider the minimization of the power lost when the reader is very close to one of the boundaries of the deployment area. Therefore, we define the power loss as follows:

$$PLoss(G) = \frac{\sum_{j=1}^k \sum_{t=1}^4 PL(r_j, B_t)}{m \cdot (2 * \rho_{Max})} \quad (10)$$

where:

$$PL(r_j, B_t) = \begin{cases} \rho_j - d_{jt} & \text{if } \rho_j - d_{jt} > 0 \\ 0 & \text{Otherwise} \end{cases} \quad (11)$$

2.2.5 Number of Deployed Readers

It is also very important to reduce the number of used readers. However, most of the related works did not consider the automatic adjustment of the number of readers. Hence, we define $DR(G)$, the ratio of the number of readers, as follows:

$$DR(G) = \frac{\sum_{j=1}^m a_j}{m} = \frac{k}{m} \quad (12)$$

where:

$$a_j = \begin{cases} 1 & \text{if the reader } r_j \text{ is selected} \\ 0 & \text{Otherwise} \end{cases} \quad (13)$$

2.2.6 Objective Function

The RNP issue addressed in this paper is a multi-objective problem consisting in simultaneously optimizing five conflicting objectives. However, in our work, we have transformed the multi-optimization problem into a single optimization problem, as most works from the literature in this field do, by adopting a weighted summation of the five objectives. Even though converting a multi-objective problem into a single-objective problem can lead to some limitations, such an approach presents several advantages:

- 1) It provides a simplicity and efficiency in solving the problem.
- 2) It allows the exploration of trade-offs between different objectives.
- 3) It facilitates decision-making by providing a single solution that can be easily compared with other algorithms.

In addition, by aggregating the five objectives into a single objective function, the users will have the ability to adjust the weights, which represent the importance given to every objective, according to their needs and their financial and technical constraints. Thus, the objective function used to evaluate the quality of solutions is defined according to the Equation (14):

$$f(G) = w_1 \cdot (1 - cov(G)) + w_2 \cdot Interf(G) + w_3 \cdot Pow(G) + w_4 \cdot PLoss(G) + w_5 \cdot DR(G) \quad (14)$$

where: $\sum_{i=1}^5 w_i = 1$ and $0 \leq w_i \leq 1$ $i \in \{1, 2, 3, 4, 5\}$.

It is evident that the described RNP issue can be considered as a minimization problem as follows:

$$\mathbf{Min}_G \left(w_1 \cdot (1 - cov(G)) + w_2 \cdot Interf(G) + w_3 \cdot Pow(G) + w_4 \cdot PLoss(G) + w_5 \cdot DR(G) \right)$$

Subject to

$$w_1 + w_2 + w_3 + w_4 + w_5 = 1$$

$$0 \leq w_i \leq 1; \forall i \in \{1, 2, 3, 4, 5\}$$

$$0 \leq x_r \leq W; \forall r \in \{1, \dots, k\}$$

$$0 \leq y_r \leq H; \forall r \in \{1, \dots, k\}$$

$$\rho_{Min} \leq \rho_r \leq \rho_{Max}; \forall r \in \{1, \dots, k\}$$

$$0 \leq \left(\sum_{r=1}^m C_{rt} \right) \leq 1; \forall t \in \{1, \dots, n\}, \forall r \in \{1, \dots, k\}.$$

3. PROPOSAL

In this paper, we address the problem of RFID network planning by using AHC algorithm. Hence, a brief description of this algorithm is first provided.

3.1 Ascendant Hierarchical Clustering Algorithm

Ascendant hierarchical clustering, also known as agglomerative hierarchical clustering, is a bottom-up clustering method widely used to group similar data together. It is a widely used clustering algorithm in data analysis and machine learning. The basic idea behind AHC is very simple. It starts by creating as many clusters as data points and then assigning each of these data points to a single cluster (each cluster contains one data point). Then, it performs an iterative merging of the closest pair of clusters in a single larger cluster. This process of merging is based on a similarity measure between clusters defined beforehand.

The key operations of AHC algorithm are straight forward:

1. Let each data point be a cluster,
2. Compute the similarity (proximity) matrix,
3. Merge the two closest clusters,
4. Update the similarity matrix,
5. Repeat steps 3 and 4 until only a single cluster remains.

Algorithm 1: AHC algorithm for RNP problem

Inputs: X : Set of input data (tags)
 X_Size : Size of the instance (number of tags)
Output: $Best_G$: Best Clustering
 G_Size : number of clusters
Variables: S_{ij} : Similarity matrix
 G : Set of Clusters

```

1 Function AHC()
2    $G \leftarrow \emptyset$ 
3   For  $i \leftarrow 1$  to  $n$  do
4      $G_i \leftarrow \{X_i\}$  //Create a cluster for each data point
5      $G \leftarrow G + G_i$  // Add  $G_i$  to the set of clusters
6     For  $j \leftarrow (i + 1)$  to  $n$  do
7        $S_{ij} \leftarrow Similarity(i, j)$ 
8        $S_{ji} \leftarrow S_{ij}$ 
9     End For

```

```

10    $F_i \leftarrow f(G)$  // using eq. (14)
11    $Best\_G \leftarrow G$ 
12    $Best\_F \leftarrow F_i$ 
13   End For
14    $G\_Size \leftarrow n$ 
15   While  $G\_Size > 1$  do
16      $(i', j') \leftarrow Argmin_{(i,j)}(Similarity(i, j))$ 
17     /** Merge clusters  $G_{i'}$  and  $G_{j'}$  into cluster  $G_k$  */
18      $G_k \leftarrow G_{i'} \cup G_{j'}$ 
19      $G \leftarrow G + G_k - G_{i'} - G_{j'}$ 
20      $G\_Size \leftarrow G\_Size - 1$ 
21     For  $i \leftarrow 1$  to  $G\_Size$  do
22        $S_{ki} \leftarrow Similarity(k, i)$ 
23        $S_{ik} \leftarrow S_{ki}$ 
24     End For
25      $F_i \leftarrow f(G)$  // using eq. (14)
26     If  $F_i < Best\_F$  then
27        $Best\_G \leftarrow G$ 
28        $Best\_F \leftarrow F_i$ 
29     End If
30   End While
31   Return  $Best\_G$ 
32End

```

3.2 AHC Algorithm Applied to RNP Issue

The described RNP issue can be considered as a clustering problem. In fact, the tags are the objects to make in clusters and every cluster represents a reader, such that the centre of gravity of the clusters defines the position of the reader and the radius of the cluster determines the communication range (power) of the reader. From this observation, we have solved the RNP problem using a clustering algorithm. In the following, a detailed description of the proposed approach to solve the considered RNP problem is given. This approach is based on hierarchical clustering. There exists two categories of hierarchical clustering: Ascendant Hierarchical Clustering (AHC) and Descendant Hierarchical Clustering (DHC). In this work, we apply the AHC method to solve the above RNP problem. The detailed pseudo-code of the AHC algorithm is shown in Algorithm 1.

To be more efficient, the solution obtained from this algorithm is then adjusted by removing the readers that are covering less than the defined threshold of the number of tags. Furthermore, the radius of some readers is adjusted to fit the maximum radius when it exceeds this value.

3.2.1 Similarity Measure

There are different ways to measure the similarity between two clusters. The most common one is the Euclidean distance. In addition, it is very important to determine a linkage criterion, which defines the way to compute the distance between two clusters based on the links between their constituent data points. Accordingly, there are several approaches to define this linkage, such as: minimum (single) linkage, maximum (complete) linkage and average linkage. In this paper, we define the distance between two clusters as the Euclidean distance between their centres of gravity. It is defined as follows:

$$S_{ij} = D(o_i, o_j) = \sqrt{\sum_{k=1}^d (x_k^i - x_k^j)^2} \quad (15)$$

where: S_{ij} is the similarity measure between clusters c_i and c_j . D is the Euclidean distance and o_i and o_j are the centers of gravity of cluster c_i and cluster c_j , respectively.

4. PERFORMANCE EVALUATION

To evaluate the performance of the proposed AHC algorithm, we have proceeded by simulation. Hence, the proposed AHC algorithm has been compared to two algorithms from the literature: PS²O [7] and

LaF [14]. PS²O [7] is commonly used for comparing new proposed RNP algorithms, while LaF algorithm [14] was proposed in 2021. Furthermore, a simulator has been developed using C++ language. This simulator includes the following features:

- Implementation of the mathematic model of the RNP problem.
- Implementation of three RNP algorithms: AHC, PSO and LaF.
- Implementation of a visualization tool.
- Implementation of dataset generator.

The simulation procedure as well as the obtained results are described in the following part.

4.1 Simulation Parameters

The parameters used during the simulation as well as their values are summarized in Table 2.

Table 2. Simulation parameters and their values.

<i>Parameter</i>	<i>Possible values</i>	<i>Default</i>
H, W	{20, 30, 40, 50, 60}	30
<i>n</i>	{50, 100, 200, 300, 400}	100
<i>m</i>	{50, 100, 200, 300, 400}	100
ρ_{Min}	1 <i>m</i>	1 <i>m</i>
ρ_{Max}	5 <i>m</i>	5 <i>m</i>
w_1	[0, 1]	0.2
w_2	[0, 1]	0.15
w_3	[0, 1]	0.05
w_4	[0, 1]	0.1
w_5	[0, 1]	0.5
p		30
T	10 000	10 000
L	40	40
S	2 <i>m</i>	2 <i>m</i>
#runs		20

The datasets used during the simulation have been generated randomly. In fact, 200 datasets representing 10 scenarios have been generated according to the parameters illustrated in Table 3. Mainly, these datasets have been generated using two tag distributions: uniform and clustered. Hence, for every scenario, 20 datasets are generated using the same parameter values.

Table 3. Parameter settings of the RNP instances.

<i>Scenario</i>	<i>Tag distrib.</i>	<i>Field</i>	<i>#Tags</i>	<i>#MaxReaders</i>
R50	Uniform	20 x 20	50	50
R100	Uniform	30 x 30	100	100
R200	Uniform	40 x 40	200	200
R300	Uniform	50 x 50	300	300
R400	Uniform	60 x 60	400	400
C50	Clustered	20 x 20	50	50
C100	Clustered	30 x 30	100	100
C200	Clustered	40 x 40	200	200
C300	Clustered	50 x 50	300	300
C400	Clustered	60 x 60	400	400

4.2 Results and Discussion

In the following part, we illustrate and describe the results obtained from simulations. Hence, Figure 1 shows the implemented simulator. The latter includes algorithm selection, parameter setting and obtained results in terms of different objectives and metrics. Figure 2 (Figure 4, respectively) illustrates the visualization of the solutions found by AHC, PSO and LaF while considering the parameters shown in Figure 1 (Figure 3, respectively).

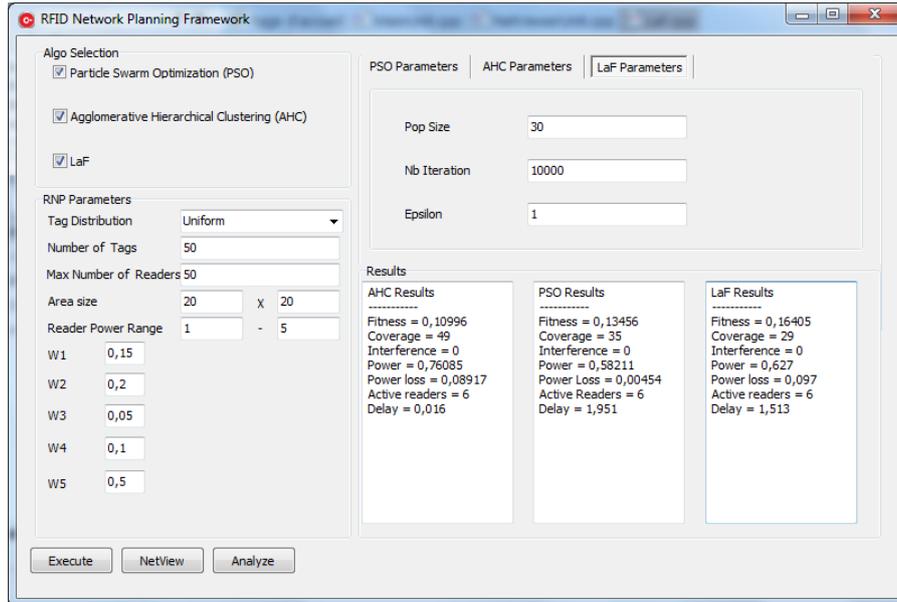


Figure 1. Parameter setting for an R50 instance.

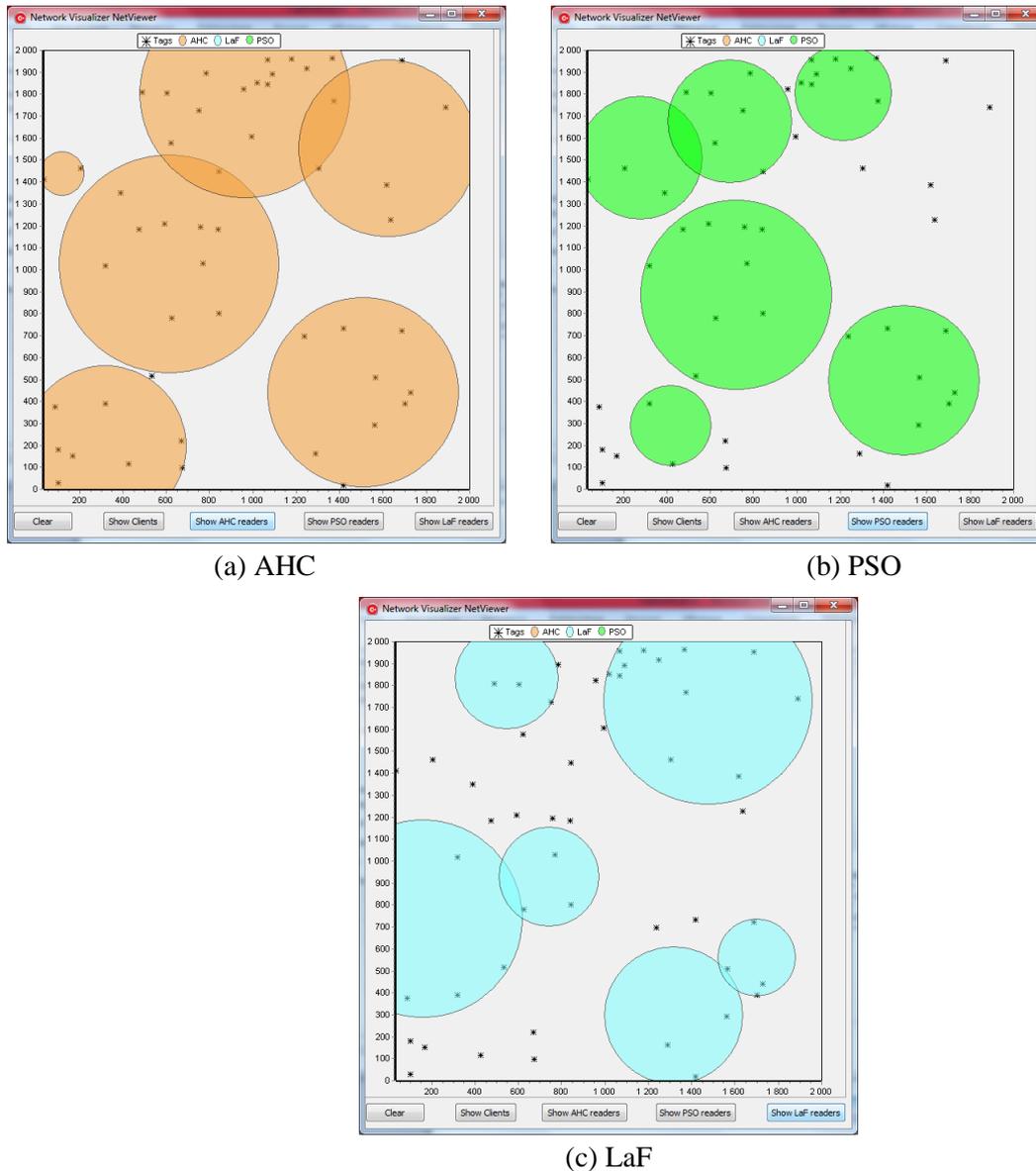


Figure 2. Results obtained by AHC, PSO and LaF for an R50 instance.

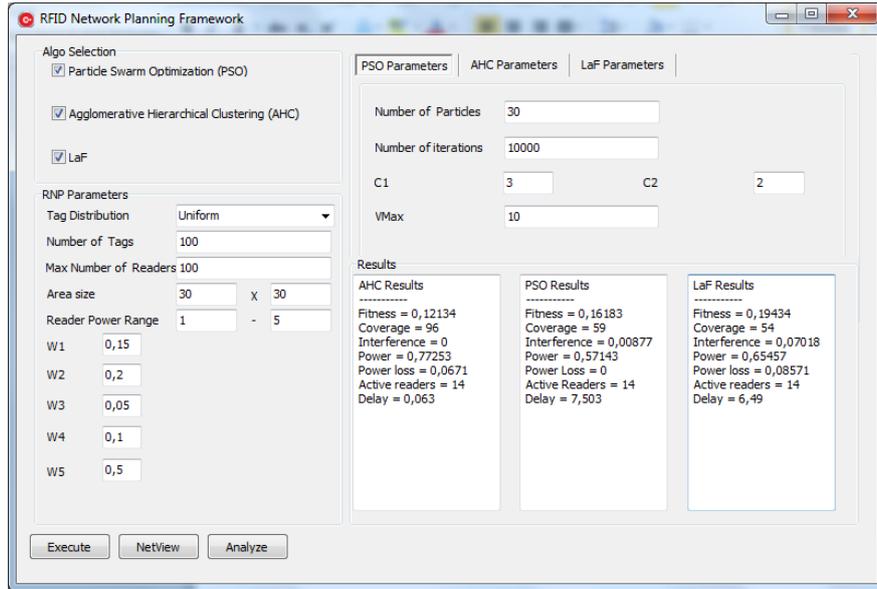
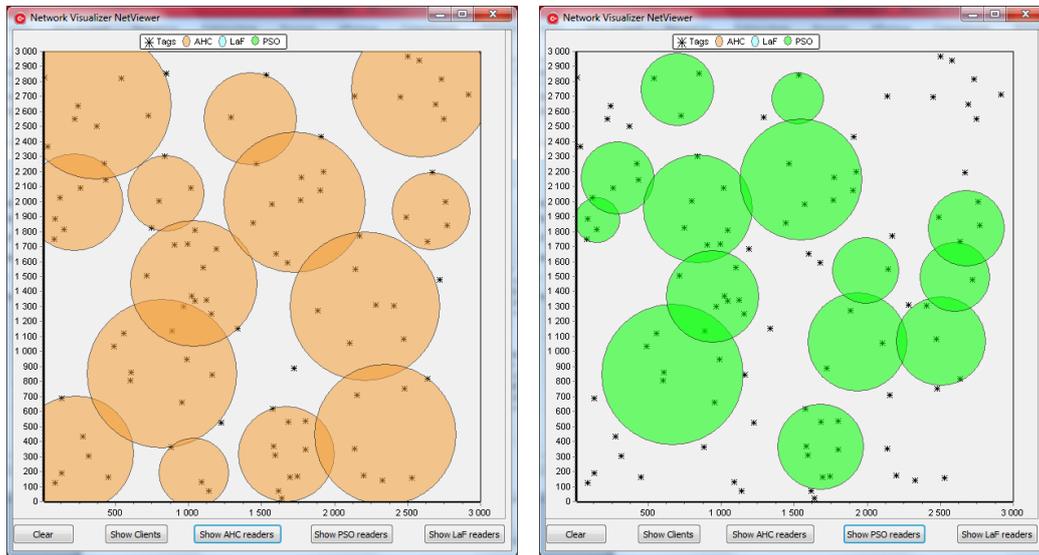
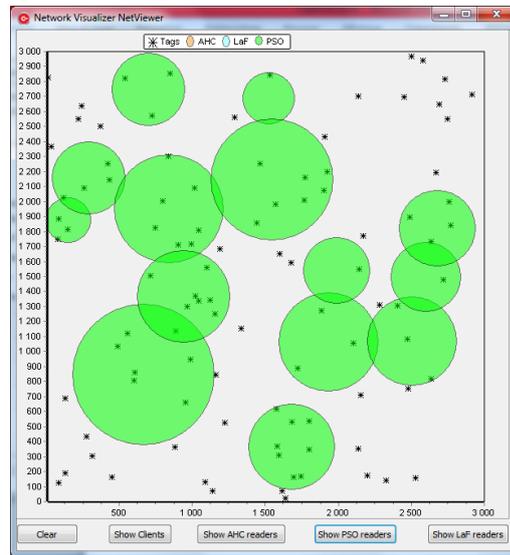


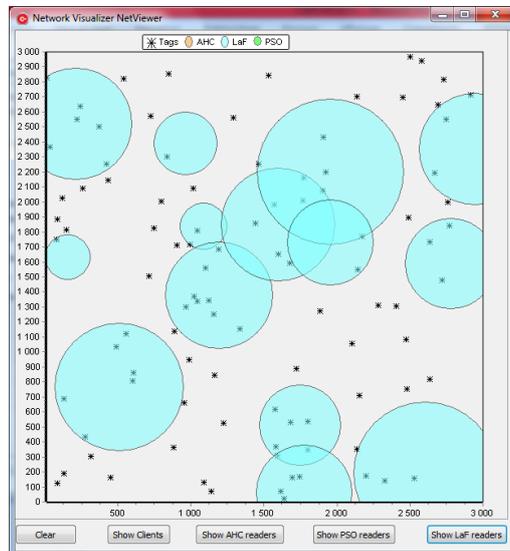
Figure 3. Parameter setting for an R100 instance.



(a) AHC



(b) PSO



(c) LaF

Figure 4. Results obtained by AHC, PSO and LaF for an R100 instance.

4.2.1 Evaluating the Impact of Varying Objective Weights

Even though our work aims at finding a good (eventually best) solution to the considered RNP problem regardless of the values of weights, it is very important to study the impact of varying the values of these weights on the quality of the solution and the performance of the system. Thus, Table 4 shows the different weight settings used to measure the impact of these settings on the performance of the system. In fact, the greatest the value of weight, the more importance is given to the corresponding objective. Therefore, weight setting WS #1 gives the same importance to all metrics, while the other settings define different ascendant orders of objective importance.

Table 4. Different weight settings.

	<i>Weight settings</i>					
	<i>WS #1</i>	<i>WS #2</i>	<i>WS #3</i>	<i>WS #4</i>	<i>WS #5</i>	<i>WS #6</i>
w_5 (<i>Selected readers</i>)	0.2	0.5	0.05	0.1	0.15	0.2
w_1 (<i>Coverage</i>)	0.2	0.2	0.5	0.05	0.1	0.15
w_2 (<i>Interference</i>)	0.2	0.15	0.2	0.5	0.05	0.1
w_3 (<i>Power</i>)	0.2	0.1	0.15	0.2	0.5	0.05
w_4 (<i>Power loss</i>)	0.2	0.05	0.1	0.15	0.2	0.5

Table 5 illustrates the results obtained when varying weight setting according to Table 4 and using the scenario R50. The results obtained represent average values of 20 randomly generated instances using the same parameter setting. Hence, these results demonstrate clearly that the behavior of the three algorithms changes according to the weight settings all algorithms adapt to the user needs. However, in practice, selected readers, coverage and interference are, generally, given more importance by setting their weights to higher values. Accordingly, the weights of these metrics are defined according to Table 1.

Table 5. Evaluating the impact of objective weights.

<i>Metric</i>	<i>Weight Setting</i>	<i>AHC</i>	<i>PSO</i>	<i>LaF</i>	<i>Metric</i>	<i>Weight setting</i>	<i>AHC</i>	<i>PSO</i>	<i>LaF</i>
Fitness	<i>WS #1</i>	0.129	0.171	0.235	Power	<i>WS #1</i>	0.3004	0.3506	0.4241
	<i>WS #2</i>	0.147	0.158	0.201		<i>WS #2</i>	0.6092	0.4344	0.5329
	<i>WS #3</i>	0.058	0.146	0.210		<i>WS #3</i>	0.2294	0.4137	0.4658
	<i>WS #4</i>	0.093	0.160	0.235		<i>WS #4</i>	0.2470	0.3148	0.4039
	<i>WS #5</i>	0.183	0.211	0.293		<i>WS #5</i>	0.2224	0.2795	0.3925
	<i>WS #6</i>	0.080	0.099	0.140		<i>WS #6</i>	0.4178	0.4065	0.4879
Coverage	<i>WS #1</i>	50	38.7	36.75	Power loss	<i>WS #1</i>	0.0170	0.0143	0.0643
	<i>WS #2</i>	50	34.1	31.5		<i>WS #2</i>	0.0613	0.0119	0.0673
	<i>WS #3</i>	50	48.4	45.9		<i>WS #3</i>	0.0116	0.0471	0.1041
	<i>WS #4</i>	50	31.4	33.85		<i>WS #4</i>	0.0141	0.0106	0.0957
	<i>WS #5</i>	50	34.85	37.8		<i>WS #5</i>	0.0103	0.0090	0.0602
	<i>WS #6</i>	50	36.05	34.1		<i>WS #6</i>	0.0247	0.0045	0.0368
Interference	<i>WS #1</i>	0	0.0804	0.1964	Selected readers	<i>WS #1</i>	16.5	16.5	16.5
	<i>WS #2</i>	0.0008	0.0128	0.0446		<i>WS #2</i>	8.05	8.05	8.05
	<i>WS #3</i>	0	0.2481	0.3606		<i>WS #3</i>	23.1	23.1	23.1
	<i>WS #4</i>	0	0.1103	0.1893		<i>WS #4</i>	20.9	20.9	20.9
	<i>WS #5</i>	0	0.2017	0.3315		<i>WS #5</i>	23.35	23.35	23.35
	<i>WS #6</i>	0.0008	0.0572	0.1221		<i>WS #6</i>	11.6	11.6	11.6

4.2.2 Fitness

The fitness value is the most important parameter that allows evaluating the performance of any given RNP algorithm. Table 6 shows the obtained fitness for the three algorithms according to different scenarios. The results shown in Table 6 demonstrate that the proposed AHC algorithm clearly outperforms the two PSO and LaF algorithms. In addition, AHC algorithm succeeded in maintaining almost the same performance (between 0.109 and 0.118 for uniform tag distribution scenarios and between 0.088 and 0.092 for clustered tag distribution scenarios). The fitness of PSO and LaF increases when increasing the number of tags. Consequently, AHC algorithm is more scalable than PSO and LaF algorithms when addressing the RNP problem.

Table 6. Obtained fitness according to different scenarios.

Scenario	AHC			PSO			LaF		
	Worst	Average	Best	Worst	Average	Best	Worst	Average	Best
R50	0.127	0.118	0.105	0.141	0.128	0.115	0.171	0.157	0.141
R100	0.124	0.117	0.109	0.158	0.144	0.126	0.184	0.170	0.151
R200	0.114	0.109	0.103	0.163	0.151	0.137	0.183	0.175	0.168
R300	0.114	0.109	0.104	0.172	0.161	0.148	0.190	0.183	0.171
R400	0.115	0.111	0.107	0.183	0.168	0.150	0.194	0.186	0.178
C50	0.106	0.089	0.065	0.120	0.096	0.069	0.140	0.123	0.103
C100	0.102	0.092	0.073	0.147	0.120	0.085	0.174	0.152	0.130
C200	0.099	0.088	0.082	0.148	0.128	0.105	0.169	0.157	0.140
C300	0.101	0.090	0.081	0.155	0.143	0.129	0.175	0.166	0.154
C400	0.099	0.091	0.087	0.164	0.150	0.130	0.188	0.174	0.166

4.2.3 Coverage

Table 7 illustrates the evolution of the number of tags covered by RFID readers according to 10 scenarios containing 20 instances each. Again, AHC algorithm outperforms PSO and LaF algorithms by covering more RFID tags. Hence, AHC algorithm covers between 96.4% and 97.9% tags, while PSO algorithm covers between 46% and 76% tags and LaF algorithm covers between 43% and 67% tags. The AHC algorithm is designed to cover all tags; however, when a reader is covering a much less number of tags, it is removed. In addition, the radius of a reader may be decreased to fit the defined threshold, which means that few tags may remain not covered.

Table 7. Obtained coverage according to different scenarios

Scenario	AHC			PSO			LaF		
	Worst	Average	Best	Worst	Average	Best	Worst	Average	Best
R50	42	48.95	50	20	38	46	17	33.8	45
R100	88	97.05	100	37	64.55	80	38	52.8	67
R200	180	193.15	200	82	118	158	72	100.85	122
R300	279	289.15	299	115	154.4	179	103	135.65	159
R400	368	386.25	396	149	190.75	245	142	173.9	216
C50	43	48.95	50	24	38	46	28	33.8	45
C100	92	97.05	100	48	64.55	80	32	52.8	67
C200	181	193.15	200	88	118	158	79	100.85	122
C300	280	289.15	299	119	154.4	179	118	135.65	159
C400	369	386.25	396	140	190.75	245	137	173.9	216

4.2.4 Interference

Interferences should also be minimized when addressing the RNP issue. Table 8 illustrates the obtained interference for the three algorithms according to 10 scenarios. This table shows, once again, that overall, our proposed algorithm is better than PSO and LaF algorithms in terms of interference. Indeed, AHC algorithm encourages merging close clusters (readers), resulting in distant clusters and reduced overlaps. Consequently, interference is minimized.

Table 8. Obtained interference value according to different scenarios.

Scenario	AHC			PSO			LaF		
	Worst	Average	Best	Worst	Average	Best	Worst	Average	Best
R50	0	0	0	0.017	0.00085	0	0.054	0.01930	0
R100	0.009	0.00045	0	0.061	0.00790	0	0.086	0.03015	0.009
R200	0.004	0.00020	0	0.054	0.01645	0	0.093	0.04395	0.013
R300	0	0	0	0.091	0.0318	0.003	0.124	0.06255	0.018
R400	0.002	0.00010	0	0.049	0.02655	0.007	0.091	0.05875	0.020
C50	0.018	0.00180	0	0	0	0	0.055	0.01010	0
C100	0.018	0.00225	0	0.009	0.00135	0	0.062	0.02195	0
C200	0.009	0.00145	0	0.009	0.00415	0	0.070	0.03025	0
C300	0.006	0.00105	0	0.042	0.01140	0	0.086	0.0399	0.006
C400	0.005	0.00085	0	0.051	0.01980	0.002	0.115	0.05445	0

4.2.5 Power

The results shown in Table 9 describe a comparison between the transmitted powers of AHC, PSO and LaF algorithms. Since we are giving more importance to the coverage in the objective function, our proposed algorithm is making more effort to maximize coverage and minimize the number of readers. To achieve this goal, clusters should be larger, such that more tags are included and fewer readers are used. For this reason, AHC algorithm needs more power than PSO and LaF algorithm.

Table 9. Obtained power according to different scenarios.

Scenario	AHC			PSO			LaF		
	Worst	Average	Best	Worst	Average	Best	Worst	Average	Best
R50	0.858	0.696	0.556	0.560	0.441	0.324	0.667	0.495	0.353
R100	0.867	0.723	0.598	0.631	0.462	0.320	0.617	0.542	0.456
R200	0.810	0.740	0.674	0.673	0.509	0.413	0.625	0.557	0.477
R300	0.792	0.729	0.658	0.561	0.516	0.442	0.651	0.571	0.477
R400	0.810	0.750	0.691	0.595	0.533	0.463	0.615	0.567	0.512
C50	0.482	0.633	0.857	0.306	0.508	0.717	0.373	0.584	0.814
C100	0.541	0.630	0.819	0.414	0.510	0.649	0.396	0.546	0.686
C200	0.562	0.658	0.792	0.443	0.536	0.681	0.441	0.594	0.710
C300	0.579	0.667	0.784	0.412	0.534	0.608	0.498	0.573	0.634
C400	0.576	0.677	0.765	0.462	0.530	0.612	0.500	0.592	0.660

4.2.6 Power Loss

Measuring the power lost outside the boundaries of the deployment area is another way to measure the efficiency of any RNP algorithm, as illustrated in Table 10. As shown in this table, PSO algorithm outperforms both AHC and LaF algorithms. Furthermore, AHC and LaF algorithms have, approximately, equivalent performances in terms of power loss.

Table 10. Obtained power loss according to different scenarios.

Scenario	AHC			PSO			LaF		
	Worst	Average	Best	Worst	Average	Best	Worst	Average	Best
R50	0.127	0.0774	0.0420	0.0350	0.0071	0	0.118	0.0602	0
R100	0.123	0.0588	0.0240	0.0120	0.0036	0	0.162	0.0566	0.013
R200	0.075	0.0477	0.0290	0.0220	0.0059	0	0.088	0.0418	0.015
R300	0.064	0.0368	0.0200	0.0100	0.0024	0	0.064	0.0353	0.013
R400	0.046	0.0328	0.0210	0.0100	0.0031	0	0.043	0.0253	0.007
C50	0.151	0.0751	0.0170	0.0550	0.0173	0	0.150	0.0737	0
C100	0.088	0.0401	0.0050	0.0520	0.0078	0	0.138	0.0465	0.011
C200	0.078	0.0422	0.0150	0.0300	0.0067	0	0.137	0.0445	0.001
C300	0.059	0.0336	0.0080	0.0120	0.0027	0	0.065	0.0343	0.013
C400	0.035	0.0253	0.0110	0.0090	0.0020	0	0.058	0.0278	0.010

4.2.7 Number of Readers

The proposed AHC algorithm is characterized by its ability to determine the adequate number of readers required to achieve good performance of the RFID system. Therefore, Figure 5 illustrates the impact of increasing the number of tags on the number of readers required to cover them. Indeed, the number of readers increases when increasing the number of tags. However, clustered scenarios require less readers than uniform scenarios. When tags are grouped in clusters, it is easy to find fewer readers to cover them. On the other hand, when tags are uniformly distributed on the deployment area, more readers are needed to cover them.

4.2.8 Execution Time

Figure 6 illustrates the execution time of all algorithms when increasing the number of tags. The most important observation from this figure is that our proposed AHC algorithm maintains, almost, the same

execution time when increasing the number of tags. It is observed that AHC algorithm gives the result in less than 4 seconds even with large instances, while PSO and LaF algorithms can run in more than 240 seconds when the number of tags exceeds 200. These results demonstrate that AHC algorithm is more scalable than PSO and LaF algorithms.

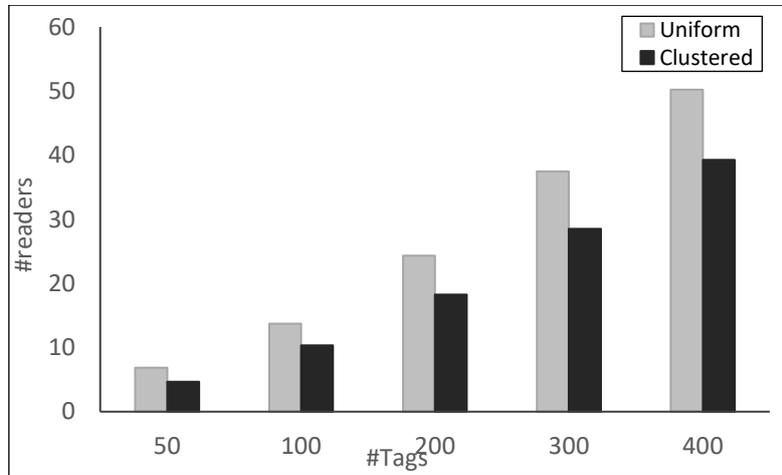


Figure 5. Number of deployed readers according to different scenarios.

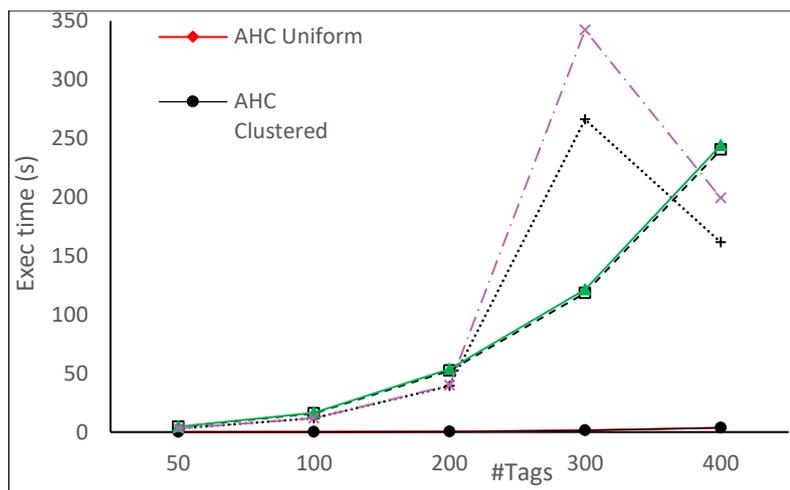


Figure 6. Execution time according to different scenarios.

5. CONCLUSION

In this paper, we have addressed the problem of RFID network planning (RNP) by considering the optimization of five parameters: coverage, interference, power, power loss and number of readers. We have considered the RNP as a clustering problem by grouping the closer tags into the same clusters. Hence, an ascendant hierarchical clustering (AHC) algorithm has been proposed to solve this issue. The results obtained have demonstrated that the proposed AHC algorithm outperforms PSO and LaF algorithms.

In the future, we expect to consider other variants of the problem, such as introducing obstacles. In addition, we expect to apply other algorithms, like Descendant Hierarchical Clustering, to solve this issue.

REFERENCES

- [1] B. Niu, E. C. Wong, Y. Chai and L. Li, "RFID Network Planning Based on MCP SO Algorithm," Proc. of the 2009 2nd Int. Symposium on Information Science and Engineering, pp. 8-12, Shanghai, China, 2009.
- [2] "RFID Market," [Online], Available: <https://www.marketsandmarkets.com/Market-Reports/rfid-market-446.html>, 2022.

- [3] C. Zhao, J. Zhu, Z. Xu, S. Chen, F. Chen and T. Wang, "Wear-free Indoor Fall Detection Based on RFID and Deep Residual Networks," International Journal of Communication Systems, vol. 36, no. 10, 2023.
- [4] C. Zhao, F. Xiong, T. Wang, Y. Wang, F. Chen and Z. Xu, "Wear-free Gesture Recognition Based on Residual Features of RFID Signals," Intelligent Data Analysis, vol. 26, no. 14, pp. 1051-1070, 2022.
- [5] Q. Guan, Y. Liu, Y. Yang and W. Yu, "Genetic Approach for Network Planning in the RFID Systems," Proc. of the 6th Int. Conf. on Intelligent Systems Design and Applications, Jian, China, 2006.
- [6] Y. Cao, J. Liu and Z. Xu, "A Hybrid Particle Swarm Optimization Algorithm for RFID Network Planning," Soft Computing, vol. 25, pp. 5747-5761, 2021.
- [7] H. Chen, Y. Zhu, K. Hu and T. Ku, "RFID Network Planning Using a Multi-swarm Optimizer," Journal of Network and Computer Applications, vol. 34, no. 13, pp. 888-901, 2011.
- [8] Z. Zhang, J. Ren, Z. Dai, B. Zhang, L. Wu and Z. Wu, "Research on RFID Network Planning Based on Directional and Omnidirectional Antenna Readers," Proc. of the Int. Conf. on Natural Computation, Fuzzy Systems and Knowledge Discovery, pp. 1192-1202, Xi'an, China, 2020.
- [9] C. Zhao, C. Wu, J. Chai, X. Wang, X. Yang, J.-M. Lee and M. J. Kim, "Decomposition-based Multi-objective Firefly Algorithm for RFID Network Planning with Uncertainty," Applied Soft Computing, vol. 55, pp. 549-564, 2017.
- [10] M. Tuba, N. Bacanin and A. Alihodzic, "Firefly Algorithm for Multi-objective RFID Network Planning Problem," Proc. of the 2014 22nd Telecom. Forum Telfor (TELFOR), pp. 95-98, Belgrade, Serbia, 2014.
- [11] J. Zhang, J. Zheng, X. Xie, Z. Lin and H. Li, "Mayfly Sparrow Search Hybrid Algorithm for RFID Network Planning," IEEE Sensors Journal, vol. 22, no. 116, pp. 16673-16686, 2022.
- [12] M. Shokouhifar, "Swarm Intelligence RFID Network Planning Using Multi-antenna Readers for Asset Tracking in Hospital Environments," Computer Networks, vol. 198, p. 108427, 2021.
- [13] M. Maimouni and B. Abou El Majd, "New Hybrid Approach Based on Artificial Neural Networks for RFID Network Planning with Random Redundant Antenna Elimination," Proc. of the 2021 3rd Int. Conf. on Transportation and Smart Technologies (TST), pp. 24-29, Tangier, Morocco, 2021.
- [14] T. Mathaba, "Multi-objective Optimal RFID Reader Deployment Using a Leaders and Followers Algorithm," Computers & Electrical Engineering, vol. 94, p. 107323, 2021.
- [15] E. G. Zahran, A. A. Arafa, H. I. Saleh and M. I. Dessouky, "A Self Learned Invasive Weed-mixed Biogeography Based Optimization Algorithm for RFID Network Planning," Wireless Networks, vol. 26, no. 16, pp. 4109-4127, 2020.
- [16] A. Jaballah and A. Meddeb, "A New Variant of Cuckoo Search Algorithm with Self Adaptive Parameters to Solve Complex RFID Network Planning Problem," Wireless Networks, vol. 25, pp. 1585-1604, 2019.
- [17] H. Chen, Y. Zhu and K. Hu, "Multi-colony Bacteria Foraging Optimization with Cell-to-cell Communication for RFID Network Planning," Applied Soft Computing, vol. 10, no. 12, pp. 539-547, 2010.
- [18] O. Franek, "Phasor Alternatives to Friis' Transmission Equation," IEEE Antennas and Wireless Propagation Letters, vol. 17, no. 11, pp. 90-93, 2017.

ملخص البحث:

يمثل تحديد الترددات الراديوية (RFID) تكنولوجيا ملائمة للعديد من سيناريوهات التطبيق. إلا أنه ينطوي على العديد من المسائل التي تحتاج للتعامل معها. ويُعدّ تخطيط شبكات تحديد الترددات الراديوية أحد التحديات التي تتناولها هذه الورقة بالدراسة، وتتمثل تلك المسألة في الموضوعة المثالية لقارئات تحديد الترددات الراديوية، ومدى الاتصال. في هذه الدراسة، جرى اعتبار المسألة المشار إليها أعلاه مسألة عنقودة؛ فقد تمّت معالجتها باستخدام خوارزمية عنقودة هرمية تصاعدية (AHC)، بحيث يمثل رأس كل عنقود إحدى القارئات. وتتميز الطريقة المقترحة هنا بأنها -بالإضافة إلى تحقيقها التخطيط المثالي لشبكات الترددات الراديوية- تعمل على تحديد العدد المثالي للقارئات اللازم لتحقيق أفضل أداء.

وقد أثبتت التجارب عبر المحاكاة أنّ الخوارزمية المستخدمة في هذه الدراسة تفوّقت على اثنتين من الخوارزميات الواردة في أدبيات الموضوع (PSO و Laf). فقد ثبت أن الخوارزمية المستخدمة في هذه الدراسة (AHC) أكثر فاعلية، وأكثر قابلية للتوسيع، وتعمل على نحوٍ أسرع عند مقارنتها بالخوارزميتين الأخرين.

TO DISTRIBUTE THE ABNORMAL EVENT USING CM-DDS ALGORITHM, WITH SECURELY USING RSA-IVDE ALGORITHM

Kande Archana and V. Kamakshi Prasad

(Received: 25-Oct.-2023, Revised: 18-Dec.-2023, Accepted: 1-Jan.-2024)

ABSTRACT

Automating the secure distribution and storage of video data addresses inefficiencies in manual processes. The emphasis on secure distributed storage aligns with the increasing importance of safeguarding video data. Traditional approaches in distributed video-data storage face limitations, like lack of scalability, data duplication, privacy and time complexity challenges in surveillance systems and suboptimal storage efficiency. This research addresses the existing limitations in distributed storage for video data by proposing a Central Moment-driven Distributed Structure (CM-DDS) and an incremental RSA-driven Incremental Video Data Encryption (RSA-IVDE) method. The framework optimizes video-data distribution across nodes, enhancing redundancy and fault tolerance for continuous data availability, even in the event of node failures. Emphasizing simplicity and accessibility, the user-friendly interface facilitates efficient management of large video datasets across various applications, including social networks, video games, surveillance and transportation networks. This comprehensive approach aims to provide a robust solution for secure distributed storage and transmission of video data while addressing the challenges identified in the existing research. The proposed approach surpasses other traditional methods with minimal decryption mean time (0.208s), minimal encryption mean time (0.182s) and minimal distribution mean time (1.620s).

KEYWORDS

Access control, Automation, Distributed storage, Scalability, Security, Video data.

1. INTRODUCTION

In today's world, the widespread use of video data has surged significantly, establishing itself as a valuable asset across various industries, particularly in security, surveillance, media and entertainment. This increasing prevalence of video data underscores the need for secure storage and efficient management practices. The exponential increase in video data poses challenges in storage and management, especially in distributed environments with complexities related to data scale, security and accessibility standards [1].

The realm of video-data management is inherently tied to the challenge of dealing with the sheer magnitude of data generated daily. This surge is fueled by technological advancements, consumer demands and the ubiquity of devices equipped with high-quality cameras [2]. As video data becomes integral to various aspects of modern life, including surveillance, filmmaking and social media, the demand for robust infrastructure has surged. Such infrastructure must not only accommodate the influx of video data, but also ensure its security, seamless accessibility and efficient organization. To address the challenges linked with video data storage, a nuanced understanding of the distributed-storage environment becomes crucial.

The inherent complexities of distributed storage systems, where data is spread across multiple nodes or locations, necessitate innovative solutions to optimize efficiency while safeguarding against security threats and ensuring timely access to critical data. This intricate balance between security, accessibility and efficiency underscores the contemporary challenges faced in managing video data within distributed-storage frameworks [3]-[4].

Turning to existing approaches in the realm of video-data management, various methodologies have emerged to tackle the challenges of secure and efficient storage in distributed environments. One of the prevalent approaches involves the use of advanced compression algorithms tailored for video data.

These algorithms play a crucial role in mitigating the storage burden by reducing the size of video files without compromising content quality. However, finding the right balance between compression ratios and video quality remains a nuanced challenge. Another significant avenue of exploration revolves around distributed-storage architectures, like cloud-based solutions. Cloud storage platforms offer scalability and flexibility, allowing organizations to adapt to the dynamic nature of video-data requirements. Additionally, cloud-based solutions often integrate robust security measures and facilitate seamless accessibility, making them increasingly popular in various sectors reliant on video data [5]-[6].

Furthermore, the integration of Artificial Intelligence (AI) [7] and Machine Learning (ML) [8] and encryption technologies [9]-[10] has marked a transformative paradigm in video-data management. AI-powered algorithms contribute to automated video analysis, content tagging and real-time monitoring, streamlining the storage process and enhancing the overall utility of video data. The multi-layered landscape of video-data management demands a comprehensive exploration of these diverse approaches to effectively navigate and address the associated complexities. Current video-data storage and management approaches face limitations in scalability, data duplication, security, access latency and time complexity. Ongoing research is needed to address these inaccuracies and develop more robust solutions capable of adapting to the challenges of video-data and distributed-storage environments. The major contributions of the research work are highlighted below.

- This research work integrates a CM-DDS for video frames, enhancing the framework's efficiency in managing and securing video data through statistical methods. CM-DDS excels by offering enhanced statistical representation, increased sensitivity to distribution nuances, adaptability to diverse datasets and robustness to outliers.
- It introduces an incremental RSA-IVDE method, ensuring enhanced security through the utilization of incremental RSA mechanisms. RSA-IVDE's incremental encryption ensures enhanced security and efficiency, adapting well to evolving data, supporting real-time encryption and facilitating efficient key management.
- It optimizes the distribution of video data across multiple nodes to enhance the framework's redundancy and fault tolerance. Strategies are implemented to ensure continuous data availability, even in the event of node failures.
- Finally, it emphasizes simplicity and accessibility in the framework's design, featuring a user-friendly interface for efficient management of large volumes of video data by diverse users and organizations.

The research paper is organized as follows. Section 2 reviews recent research, identifying the limitations of foundational methods for data distribution and cryptography. Section 3 proposes solutions, including an automated framework for secure distributed storage. Section 4 presents results and discussions on the framework's performance. Section 5 concludes the research paper, emphasizing its contributions and suggesting future research.

2. RECENT RESEARCH REVIEWS

After establishing the research context and introducing the baseline methods, the subsequent section delves into the latest enhancements made beyond these initial approaches.

In recent years, there has been significant research interest in the field of distributed storage for video data, with several works addressing the challenges in this domain. This sector postured numerous concerns that require attention. Yu et al. [11] have introduced two distinct architectures for the decentralized storage of surveillance video data; namely, "Centipede: Leveraging the Distributed Camera Crowd for Cooperative Video Data Storage" and "A Distributed Storage System for Robust, Privacy-Preserving Surveillance Cameras." The first work focused on a system that collaboratively stored and managed video data, leveraging the crowdsourcing capabilities of distributed cameras. It employed a hierarchical storage structure and data redundancy to ensure trustworthiness and scalability. The second work advocated for a decentralized storage system to protect individuals' constitutionally protected right to privacy when using surveillance cameras. It ensured the confidentiality of camera positions and operator identities while providing a secure means for data transmission. Wu et al. [12] proposed HuaVideo as a method to deliver secure, scalable and compatible HTML5 video content. This hybrid design combines client-server and peer-to-peer models, incorporating a novel access control

method to restrict video-content access to authorized users. Mar et al. [13] presented a private cloud storage system addressing privacy and security issues in cloud-based storage. The proposed system utilized encryption and a decentralized architecture to enhance fault tolerance and scalability. Ma et al. [14] suggested dynamic video switching for a distributed crane-control system, emphasizing bandwidth and resource utilization issues. The developed method adapted to workload changes and resource availability variations. The literature review provided an overview of recent research on secure data storage and transmission for video, spanning various applications. Ambani and Atkotiya [15] introduced a robust data-mining technique, ensuring user privacy in social networks. Besancon et al. [16] explored blockchain interoperability to enhance data flow in video games. Bálint [17] proposed decentralized methods for video storage based on blockchain technology. Du and Li [18] employed visual cryptography for privacy protection in video surveillance. Mishra and Mukhopadhyay [19] suggested a strategy for secure video storage and transmission involving encryption and compression. Jacob and Rekha [20] presented a secure file-sharing method using erasure correction code and deduplication to ensure accessibility and reduce storage space. Fitwi et al. [21] proposed a blockchain-based privacy-protection system for intelligent surveillance, prioritizing a decentralized approach. Chen et al. [22] introduced cooperative caching for popular videos among mobile users to enhance service quality and reduce network congestion. Ling et al. [23] developed a distributed multilevel storage system for visual surveillance in intelligent transportation networks, offering scalability and efficiency. Harnal and Chauhan [24] examined covert cryptography approaches for multimedia cloud-storage systems, evaluating safety, performance and computational complexity. Khan et al. [25] introduced the blockchain resource constraints in the Internet of Medical Things (BIoMT) consortium architecture, incorporating blockchain hyperledger fabric for improved security, integrity and transparency in health-related transactions.

The reviewed papers provide diverse solutions for secure data exchange, privacy preservation and effective multimedia data storage in applications like surveillance, transportation networks and cloud environments. These solutions are presented in various forms throughout the papers. The subsequent part addresses the mathematical presentation of existing-system issues.

2.1 Foundational Method for Data Distribution

Following the establishment of the research context in the previous part, this part delves into the baseline method for data distribution. Assuming that the total video data, represented as $V[]$, comprises n video frames, each frame denoted as f_i , where $i = 1, 2, \dots, n$, this is expressed in Equation (1) as follows:

$$V[] = \langle f_1, f_2, \dots, f_n \rangle \quad (1)$$

Moreover, the set of frames, denoted as $F[]$, represents segmented image frames $f_p s$ from the entire video data based on time t . This formulation is expressed in Equation (2) as follows:

$$F[] = \frac{d}{dt} V[], t = f_p s \quad (2)$$

Additionally, presuming that the assembly of storage clusters, denoted as $S[]$, accommodates multiple clusters with each cluster represented as S_i , where $i = 1, 2, \dots, k$. Thus, for a total of k clusters, this is expressed in Equation (3) as:

$$S[] = \langle S_1, S_2, \dots, S_k \rangle \quad (3)$$

Ultimately, assuming that the arbitrary function λ is a first-come, first-served basis function for assigning frame sequences to storage clusters, this is expressed in Equation (4) as:

$$\lambda\{F[] :: S[]\} \rightarrow \{f_i \rightarrow S_i\} \quad (4)$$

Therefore, according to the baseline method, the ultimate allocation of $f_i \rightarrow S_i$ is achieved through the distribution key. In the subsequent sections of this study, the recent enhancements made to the baseline methods are delved into.

2.2 Foundational Method for Cryptography

This sub-section provides an overview of the traditional cryptography method. Assuming that a given

frame, f_x is encrypted or decrypted using a specified key referred to as "key," the resulting encrypted pixel matrix, $P_x[][]$ is expressed in Equation (5). In this sub-section, the traditional cryptography method is furnished.

$$P_x[][] = \sum_{i=1}^n \sum_{j=1}^m \{f_x[i][j] \oplus Key\} \quad (5)$$

Therefore, $P_x[][]$ represents the ultimate encrypted pixel matrix. By applying the reverse operation to the encrypted data, the pixel matrix is subsequently decrypted. In the following sub-sections, the recent advancements of the baseline methods are delved into.

2.3 Problem Formulation – Mathematical Model

After realizing the baseline methods and the current developments over the baseline techniques, in this sub-section of the work, the persistent research problems are outlined. The existing research in distributed storage for video data demonstrates a growing concern for security and privacy, leading to various proposed architectures and methods. However, these solutions exhibit limitations and gaps that need addressing. Challenges include the need for trustworthy and scalable frameworks, ensuring privacy in surveillance systems and optimizing storage efficiency for multimedia content. Despite the multitude of proposed techniques, a comprehensive analysis and comparison of their effectiveness based on safety, performance and computational complexity are lacking. Additionally, there is a need for validation and accuracy assessment of established systems using diverse datasets. This research is motivated by these gaps by providing a robust and efficient solution for secure distributed storage and transmission of video data, considering the diverse applications in social networks, video games, surveillance and transportation networks.

Moving forward, let's first address the data-duplication issue. Assuming that the frame collection consists of two distinct frames, f_1 and f_2 ; furthermore, when considering two separate storage clusters labeled S_1 and S_2 , this results in the distribution possibilities shown in Equations (6), (7) and (8):

$$f_1 \rightarrow S_1 \quad (6)$$

and

$$f_2 \rightarrow S_2 \quad (7)$$

or,

$$f_1, f_2 \rightarrow S_1, S_2 \quad (8)$$

However, it is evident that the redundancy resulting from these distributions directly contributes to a higher space complexity. Secondly, it is assumed that the total time complexity for the retrieval operation of video data is T_1 , formulated for n number of frames and k number of storage clusters, as denoted in Equation (9):

$$T_1 = n * k \quad (9)$$

This is reformulated by Equation (10),

$$T_1(n) = n^2 \mid n \approx k \quad (10)$$

Video-data encryption safeguards against unauthorized access, but presents challenges like computational complexity, compatibility issues and key management concerns. Sharing encryption keys between devices is difficult, leading to higher costs and performance overhead. Despite encryption, vulnerability to decryption attacks remains a risk.

Therefore, it is assumed that the time complexity O for distributed video data is T_2 , formulated for $(n1 * m1)$ number of pixels per frame, a total of n frames and k number of distributions, then Equations (11) and (12) are formulated as:

$$T_2 = \{(n1 * m1) * n\} * k \quad (11)$$

or,

$$T_2(n) = O(n^4) \mid n1 \approx m1 \approx n \approx k \quad (12)$$

Further, the proposed solutions are discussed in the next section of this work.

3. PROPOSED SOLUTIONS

After thoroughly examining the current advancements in research and identifying the limitations of traditional systems, this section presents the proposed solutions. The first method involves a hierarchical tree-storage structure driven by central moments for frames with distribution. The second method is an incremental RSA-driven approach for encrypting video data.

The proposed system employs a fundamental technique for image identification; namely, central-moment technique. Central moments are statistical measures widely utilized in image processing and computer vision to characterize the spatial distribution of pixel intensities within an image. Specifically, central moments offer a means to quantify the distribution of pixel intensities relative to the mean intensity of an image. These moments facilitate the computation of various image features, including the image centroid orientation and scale, all of which are essential for numerous applications like object recognition, tracking and image registration. Therefore, a thorough understanding of central moments and their properties are crucial for the effective analysis and manipulation of images across various domains.

Assuming that the moments' array $M[][]$ for the entire pixel matrix collection is represented by $P_x[][]$, the regions for segmented objects need to be initially separated from the rest of the image and the intensity I is expressed in Equation (13) as follows.

$$I = RGB\{P_x[][]\} \quad (13)$$

Further, the central moment of the frames is formulated by Equation (14),

$$M[][] = \sum_{i=1}^{m1} \sum_{j=1}^{m1} \{ (P_x[i][j] - Mean(P_x[i][j]))^{n1} * (P_x[i][j] - Mean(P_x[i][j]))^{m1} \} * I \quad (14)$$

Moreover, these moment values are adequate for uniquely identifying the frames and they are employed to construct the tree-storage structure as a distributed-storage system. Assuming that the storage tree $T[]$, is built using the arbitrary function τ , the relationship is expressed by Equation (15).

$$T[] = \tau\{M[][]\} \quad (15)$$

Assuming that the time complexity for data retrieval is T_3 , this is formulated by Equations (16) and (17) as follows:

$$T_3 = n + \frac{n}{2} + \frac{n}{4} + \frac{n}{8} + \dots + 1 \quad (16)$$

or,

$$T_3(n) = \log_2(n) \quad (17)$$

Secondly, the incremental encryption for video data is presented. Assuming two frames denoted as f_1 and f_2 and utilizing the arbitrary function Δ , the difference $D[]$, between these two frames in terms of data is calculated by Equation (18):

$$D[] = \Delta\{f_1, f_2\} \quad (18)$$

Further, the encrypted data, $E[]$, is built as in Equations (19) and (20),

$$E[] = RSA\{f_i\} + RSA\{D[i+1]\} \quad (19)$$

Naturally,

$$Size\{f_{i+1}\} \gg Size\{D[i+1]\} \quad (20)$$

Therefore, the improvements are easily achieved as f_{i+1} and $D[i+1]$. Furthermore, building upon the proposed methods, the upcoming sub-section establishes the proposed algorithms.

3.1 Proposed Algorithms

Following a comprehensive understanding of the proposed methods, this sub-section introduces the proposed algorithms.

Algorithm - I: CM-DDS Algorithm
Input: Video data as $V[]$ Output: Tree Structure as $T[]$
<p>Process:</p> <ol style="list-style-type: none"> 1. Read in the video data as $V[]$ 2. Calculate the overall frames in the video. 3. For each frame as $V[i]$ <ol style="list-style-type: none"> 1. Starting from the root node, each level of the tree will have twice as many nodes as the level above it. 2. The height of the tree will be the minimum integer value that satisfies $2^h \geq$ total number of frames. 3. Initialize a binary tree data structure with the appropriate number of nodes for the given height. 4. The nodes of the tree should be numbered in a left-to-right, top-to-bottom order. 4. For each frame as $V[j]$: <ol style="list-style-type: none"> 1. The first frame will be stored in the root node. 2. For each subsequent frame, traverse down the tree from the root node to the appropriate child node based on the frame index. 3. Store the frame in the corresponding node. 5. Return the Tree Structure as $T[]$
Algorithm - II: RSA-IVDE Algorithm
Input: Tree Structure as $T[]$ Output: Encrypted data as $TE[]$
<p>Process:</p> <ol style="list-style-type: none"> 1. Select two large prime numbers, p and q 2. Compute $n=p*q$ 3. Compute $I(n)=(p-1)*(q-1)$ <p>encrypt ()</p> <ol style="list-style-type: none"> 4. Select an encryption exponent, e, that is relatively prime to $I(n)$ 5. Calculate the decryption exponent d 6. Convert the video data into a series of integers that represent the bytes of the video. 7. Divide the video data into blocks of a fixed size. 8. For each block <ol style="list-style-type: none"> 1. Convert it to a single integer using a fixed-length encoding scheme. 2. Encrypt each integer using the RSA algorithm: 3. Raise the integer to the power of the encryption exponent e, modulo n 4. Store the encrypted integers in a file or stream as $TE[]$ <p>decrypt ()</p> <ol style="list-style-type: none"> 9. Read the encrypted integers from the file or stream. 10. Decrypt each integer using the RSA algorithm 11. Raise the integer to the power of the decryption exponent d, modulo n 12. Convert each decrypted integer back to a block of bytes using the fixed-length decoding scheme. 13. Concatenate the decrypted blocks to reconstruct the original video data.

Firstly, the CM-DDS algorithm [26] is presented. Central moments, a set of statistical measures, offer insights into the distribution of pixel intensities in an image. They find application in image processing and computer vision, extracting valuable features like object size, shape and orientation. Central moments are particularly advantageous due to their invariance to translation, remaining consistent even if the image undergoes shifts or translations. The central moment of an image is defined as the expected value of a function of pixel intensities, with the function centred around the mean intensity of the image. Secondly, the RSA-IVDE algorithm [27] is introduced.

The RSA algorithm is a widely utilized public-key encryption method known for its high security and efficiency. It is commonly employed in applications requiring secure data transmission, including the encryption of image data. This article explores the application of the RSA algorithm for encrypting image data and outlines the steps involved in the encryption process (see Figure 1). The RSA algorithm employs two keys, a public key for encryption and a private key for decryption. The public key is distributed to anyone needing to send encrypted messages, while the private key remains confidential with the recipient.

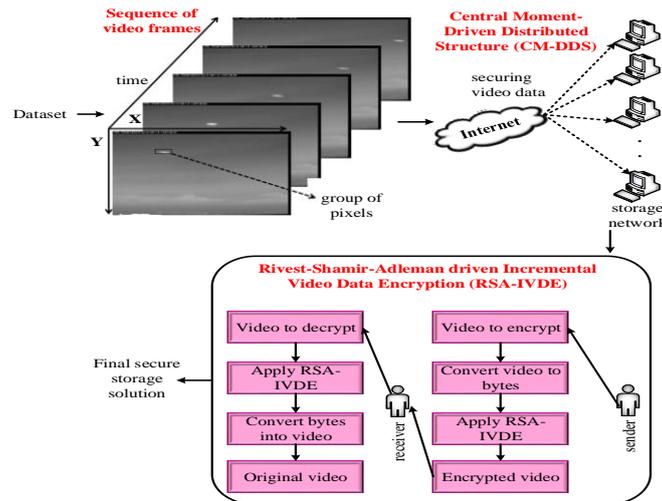


Figure 1. Proposed architecture.

To encrypt image data using RSA, the initial step involves converting the image data into a format compatible with the RSA algorithm. A common approach is to transform the image data into a series of binary values, where each pixel is represented by a binary value. This conversion is achieved through various techniques, like grayscale conversion or binary thresholding. Furthermore, the results obtained will be discussed in the upcoming section.

4. RESULTS AND DISCUSSION

This proposed work develops algorithms for two primary purposes: data distribution and data cryptography. The UCF crime dataset [28] comprises a total of 1900 samples with 7247 average number of frames and 8 number of categories. Only 15 samples are listed for illustrative purposes. Subsequently, the formations of clusters and the read/write speed for each cluster are observed and presented in Table 1.

Data distribution involves the procedure of storing and transmitting data across multiple devices or locations to guarantee availability, reliability and performance. In a distributed system, data is replicated and dispersed among multiple nodes, enhancing fault tolerance, load balancing and scalability. Data distribution is crucial for various contemporary applications, including cloud computing, big data analytics and Internet of Things (IoT) devices.

Table 1. Data-distribution analysis.

Trial Seq.	Number of Clusters (#)	Read Speed (Mbps)	Write Speed (Mbps)
1	56	148.43	137.24
2	54	109.88	99.33
3	44	110.34	143.12
4	48	106.77	88.92
5	55	91.67	126.79
6	45	131.44	112.52
7	53	96.72	141.52
8	56	122.53	114.46
9	50	133.26	141.5
10	54	140.38	97.6
11	50	126.34	125.45
12	52	127.81	106.63
13	56	101.58	121.91
14	46	127.82	104.65
15	54	105.48	91.22

Distributing data across multiple nodes is an intricate process that involves addressing issues like data consistency, synchronization and security. To achieve efficient data distribution, various techniques like replication, sharding and partitioning are employed. Replication entails copying data to multiple nodes

to ensure redundancy and availability, while sharding and partitioning involve dividing data into smaller subsets and distributing them across different nodes. Efficient data distribution is essential for modern businesses and organizations that need to manage and process large volumes of data quickly and effectively. Proper data distribution enables organizations to achieve better performance, scalability and reliability, ensuring that data is available when and where needed.

The mean read speed for the entire dataset is 120 Mbps and the write speed is 110 Mbps. The results are also illustrated diagrammatically (see Figures 2 and 3). Additionally, the formed clusters undergo testing for data loss before and after distribution, with the recorded observations presented in Table 2.

Data Loss Prevention (DLP) is a set of strategies and technologies aimed at safeguarding sensitive data from loss, theft or misuse. DLP solutions typically involve identifying, classifying and monitoring sensitive data as it moves across the network, applying controls and policies to prevent unauthorized access or exfiltration.

Effective DLP helps organizations mitigate risks related to data breaches, compliance violations and reputational damage. By implementing DLP measures, organizations reduce the likelihood of data-loss incidents, minimize their impact and maintain customer trust and regulatory compliance. Common DLP strategies include encryption, access controls, activity monitoring, data classification and incident response planning. However, DLP is not a one-size-fits-all solution and organizations should tailor their approach to their specific needs and risk profile. A comprehensive DLP program requires collaboration across multiple stakeholders, including IT, security, legal and business teams.

Thirdly, the time taken for data distribution is recorded for further analysis (see Table 3). The mean data distribution time is 1.239 seconds for the entire dataset. Distributing data across multiple clusters is a crucial technique for storing and processing large volumes of data efficiently. By distributing data across multiple clusters, organizations improve performance, increase scalability and enhance data availability and reliability. The process of data distribution involves dividing large datasets into smaller subsets and distributing them across multiple clusters in a way that maximizes performance and minimizes data duplication. This is achieved through various methods like shredding, replication and partitioning.

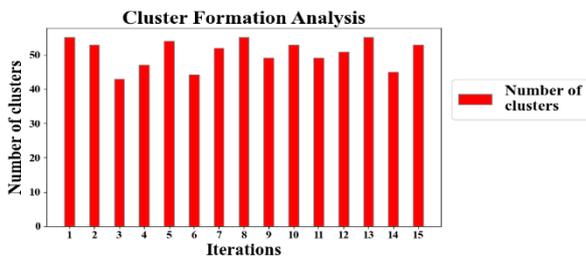


Figure 2. Cluster-formation analysis.

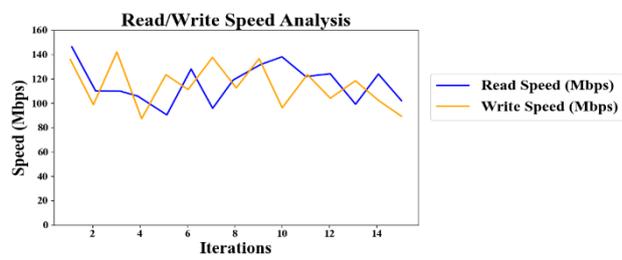


Figure 3. Read/write speed analysis.

Shredding entails breaking down datasets into smaller, more manageable subsets and distributing them across multiple clusters. Replication involves generating duplicates of datasets and storing them across multiple clusters to improve data availability and reliability. Partitioning involves segmenting datasets into smaller subsets and distributing them across multiple clusters based on specific criteria like data type, geography or user access.

Table 2. Data-distribution analysis.

Trial Seq.	Number of Clusters (#)	Data Loss (%)	Trial Seq.	Number of Clusters (#)	Data Loss (%)
1	56	0	9	50	0
2	54	0	10	54	0
3	44	0	11	50	0
4	48	0	12	52	0
5	55	0	13	56	0
6	45	0	14	46	0
7	53	0	15	54	0
8	56	0			

In summary, distributing data over multiple clusters allows organizations to harness the advantages of distributed-computing and big-data technologies, facilitating high-performance data processing, analysis and storage. The results are also visually presented in Figure 4.

Furthermore, the results derived from the cryptographic process are observed and documented in the provided details, as presented in Table 3.

Table 3. Data cryptographic analysis.

Trial Seq.	Key Length	Encryption Time (Sec)	Decryption Time (Sec)	Data Distribution Time (Sec)
1	44	0.126	0.147	1.776
2	44	0.123	0.144	1.973
3	44	0.192	0.213	1.494
4	44	0.121	0.140	1.38
5	44	0.180	0.216	1.971
6	44	0.207	0.228	1.723
7	44	0.184	0.206	1.75
8	44	0.208	0.246	1.29
9	44	0.221	0.245	1.143
10	44	0.218	0.255	1.432
11	44	0.146	0.162	1.407
12	44	0.195	0.226	1.733
13	44	0.185	0.209	1.701
14	44	0.208	0.229	1.858
15	44	0.210	0.248	1.675

For the entire dataset, the mean encryption time is 0.182 seconds and the mean decryption time is 0.208 seconds. Data encryption and decryption are crucial processes employed to safeguard sensitive information from unauthorized access and theft. Encryption transforms plain text or other forms of data into a coded format, readable only by authorized users possessing the key or password needed for decryption. Decryption, on the other hand, reverses the encryption process, restoring the data to its original form for comprehension. Data encryption ensures the confidentiality, integrity and authenticity of information by rendering it unreadable to anyone lacking access to the decryption key. It is widely utilized in industries like finance, healthcare and government, where data privacy is of paramount importance.

Two primary categories of encryption exist: symmetric encryption and asymmetric encryption. Symmetric encryption employs a single key for both encryption and decryption, while asymmetric encryption uses distinct keys for these operations. While asymmetric encryption provides enhanced security, it also demands more computational resources compared to symmetric encryption. Decryption necessitates the use of the same key or password employed during encryption to decode the data. When a user inputs the correct password or key, the encrypted data is transformed back into its original form, allowing the user to comprehend the information. The result is observed graphically by Figure 5.

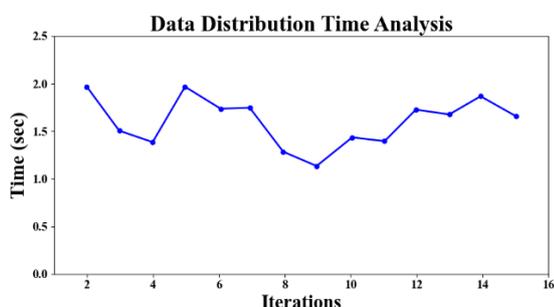


Figure 4. Data-distribution time analysis.

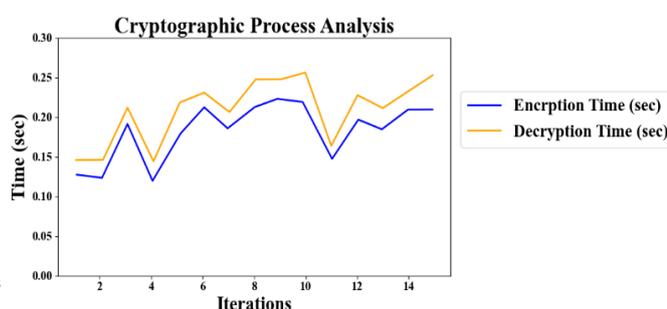


Figure 5. Encryption and decryption-time analysis.

Moreover, the proposed methodology has undergone rigorous testing, employing a diverse set of alternative datasets and the corresponding outcomes are meticulously detailed in Table 4.

Furthermore, a comparative analysis is conducted in the subsequent sub-section of this work, juxtaposing the acquired results with findings from parallel research endeavors.

Table 4. Test results on various datasets.

Dataset Name	Data Distribution Mean Time (Sec)	Encryption Mean Time (Sec)	Decryption Mean Time (Sec)
UMN [29]	1.776	0.192	0.213
UCSD [29]	1.75	0.121	0.140
Ped – 1 [29]	1.29	0.180	0.216
Ped – 2 [29]	1.143	0.207	0.228
BOSS [30]	1.432	0.184	0.206
Proposed (UCF crime)	1.407	0.208	0.246

4.1 Comparative Analysis

After conducting thorough analyses of the obtained results, this sub-section of the paper involves a comparison between the achieved outcomes and the corresponding benchmarked results, as illustrated in Table 5.

Table 5. Comparative analysis.

Ref.	Proposed Algorithm Name	No. of Segments (Mean) (#)	Mean Time for Distribution (Sec.)	Data Loss Percentage (%)	Mean Time for Encryption (Sec)	Mean Time for Decryption (Sec)
Yu et al. [11]	Centipede	34	6.11	7	3.293	4.734
Wu et al. [13]	HuaVideo	40	8.617	5	2.744	4.708
Mar et al. [14]	Secure personal cloud storage	33	6.998	5	2.936	2.339
Ambani et al. [16]	Data Mining Techniques	40	7.764	9	2.378	5.129
Besançon et al. [17]	Blockchain Interoperability	33	5.218	6	2.309	4.895
Jacob et al. [21]	Erasure	23	6.686	6	4.62	3.227
Proposed Method	Central Moment, Tree Storage Structure, RSA and Incremental backups	51	1.620	0	0.182	0.208

Thus, it becomes evident that the proposed method has surpassed a majority of the well-known benchmarked techniques in all aspects.

4.2 Case Study

Study 1: Data-distribution Analysis

In this comprehensive case study, the examination delves into the impact of different data distribution strategies—shredding, replication and partitioning—on the performance of a system managing large volumes of video data. The analysis involves six distinct scenarios: (1) lacking shredding, (2) devoid of replication, (3) without partitioning, (4) in the absence of both shredding and replication, (5) excluding both shredding and partitioning and (6) lacking both replication and partitioning. Each scenario explores how the absence of a specific distribution strategy affects data-management efficiency, fault tolerance and overall system performance. This study provides valuable insights into the nuanced interplay between these strategies and their role in optimizing data distribution for enhanced system robustness and reliability.

Scenario 1: Without Shredding

Intention: In this scenario, the CM-DDS framework is implemented without the use of shredding, which involves breaking down large datasets into smaller subsets.

Remark: While CM-DDS provides enhanced statistical representation and adaptability, the absence of

shredding results in challenges related to efficient handling of extremely large video datasets. The framework experiences limitations in terms of processing speed and resource utilization.

Scenario 2: Without Replication

Intention: Replication, which involves creating copies of data, is excluded from the CM-DDS framework.

Remark: Without replication, the redundancy and fault tolerance of the system are compromised. In the event of node failures, the availability of video data could be at risk, leading to potential interruptions in data access and processing.

Scenario 3: Without Partitioning

Intention: Partitioning, the division of data into smaller subsets, is not employed in this scenario.

Remark: The absence of partitioning might lead to suboptimal distribution of video data across nodes. This could result in uneven workloads and might affect the overall efficiency of the system, especially when dealing with diverse datasets.

Scenario 4: Without Shredding and Replication

Intention: This scenario excludes both shredding and replication.

Remark: The absence of shredding and replication results in a less robust system. Without shredding, large datasets pose processing challenges and without replication, the system lacks the necessary redundancy to ensure continuous data availability.

Scenario 5: Without Shredding and Partitioning

Intention: This scenario excludes both shredding and partitioning.

Remark: The lack of shredding and partitioning leads to difficulties in handling large volumes of video data efficiently. The absence of partitioning results in suboptimal data distribution across nodes, impacting the overall performance of the CM-DDS framework.

Scenario 6: Without Replication and Partitioning

Intention: Replication and partitioning are both excluded from the CM-DDS framework in this scenario.

Remark: Without replication, the system lacks the necessary redundancy and without partitioning, the distribution of video data is not optimized. This could lead to challenges in fault tolerance, workload distribution and overall system performance. The data-distribution time for different scenarios are depicted in Table 6.

Table 6. Data-distribution time in different scenarios.

Trial seq.	Data-distribution time (s)					
	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6
1	2.012	2.776	2.899	2.825	2.812	2.839
2	2.189	2.973	2.075	2.001	2.003	2.018
3	2.805	2.494	2.612	2.535	2.525	2.554
4	2.634	2.38	2.507	2.424	2.418	2.446
5	2.175	2.971	2.059	2.996	2.997	2.017
6	2.962	2.723	2.828	2.779	2.776	2.792
7	2.989	2.75	2.857	2.806	2.804	2.819
8	2.511	2.29	2.401	2.349	2.347	2.365
9	2.364	2.143	2.255	2.202	2.198	2.215
10	2.653	2.432	2.544	2.485	2.472	2.493
11	2.628	2.407	2.519	2.461	2.448	2.467
12	2.964	2.733	2.855	2.796	2.782	2.803
13	2.932	2.701	2.823	2.746	2.751	2.772
14	2.089	2.858	2.98	2.992	2.908	2.932
15	2.906	2.675	12798	2.738	2.726	2.748

These scenarios represent variations in the data-distribution times based on the absence of specific strategies (shredding, replication, partitioning). The values are indicative and vary based on the specific characteristics of each scenario.

Study 2: Key-length Analysis

This comprehensive case study investigates the influence of varying key lengths in a trial sequence 15, on the encryption and decryption processes of a cryptographic system. The analysis encompasses fifteen distinct trials, each with a different key length within the specified range. Table 7 explains the encryption and decryption times with respect to key length.

Table 7. Encryption and decryption times.

Trial seq.	Key length	Encryption time (s)	Decryption time (s)
15	30	0.255	0.277
15	32	0.250	0.272
15	34	0.345	0.269
15	36	0.340	0.264
15	38	0.340	0.262
15	41	0.335	0.258
15	42	0.330	0.254
15	45	0.325	0.349
15	47	0.325	0.348
15	50	0.320	0.343

The encryption and decryption times corresponding to 15-trial sequence are meticulously recorded, shedding light on the performance characteristics and computational efficiency associated with different key lengths. This study aims to discern patterns, trends and trade-offs in the cryptographic processes as the key length undergoes variation, providing valuable insights into optimizing the balance between security and computational overhead in cryptographic applications.

5. CONCLUSION

The critical gaps in existing distributed-storage solutions for video data are addressed by this research. Innovative contributions addressed challenges, enhancing video-data management through CM-DDS integration for improved representation and adaptability. The introduction of RSA-IVDE ensured heightened security and efficiency with incremental encryption, adaptable to evolving data and supporting real-time encryption. The optimization of video-data distribution across nodes improved redundancy and fault tolerance, ensuring continuous data availability. The emphasis on simplicity and accessibility in the framework's design provided an efficient solution for managing large video datasets across diverse applications. Overall, a comprehensive and robust approach to secure distributed storage and transmission of video data was contributed by this research, addressing current limitations and meeting the varied needs of users in social networks, video games, surveillance and transportation networks. The proposed approach excels in speed and efficiency, with minimal decryption (0.208s), encryption (0.182s) and distribution (1.620s) times, marking a significant advancement compared to traditional methods. These results underscore its potential to outperform existing approaches in data processing and distribution. Future research focuses on collaborating with environmental experts and integrating eco-friendly practices to enhance the ecological impact of secure distributed-storage systems, ensuring their sustainability for long-term use. This includes exploring energy-efficient strategies, incorporating renewable-energy sources and implementing environmentally conscious design principles to minimize the environmental footprint of such systems.

REFERENCES

- [1] D. Dhingra and M. Dua, "Medical Video Encryption Using Novel 2D Cosine-sine Map and Dynamic DNA Coding," *Medical & Biological Engineering & Computing*, vol.62, pp. 237-255, 2023.
- [2] S. Das, A. Adhikary, A. A. Laghari and S. Mitra, "Eldo-care: EEG with Kinect Sensor Based Telehealthcare for the Disabled and the Elderly," *Neuroscience Informatics*, vol. 3, no. 2, p. 100130, 2023.
- [3] N. Cecchinato, A. Toma, C. Drioli, G. Oliva, G. Sechi, G.L. Foresti, "Secure Real-time Multimedia Data Transmission from Low-cost UAVs with a Lightweight AES Encryption," *IEEE Communications*

- Magazine, vol. 61, no. 5, pp. 160–165, DOI:10.1109/mcom.001.2200611, 2023.
- [4] H. N. Khan, A. Das and A. Chaudhuri, "Unique Video Encryption Technique Intended for Smart City Application," *IETE Journal of Research*, vol. 69, no. 9, pp. 5830–5839, 2023.
- [5] P. Saini and K. Kumar, "S-method: Secure Multimedia Encryption Technique in Cloud Environment," *Multimedia Tools and Applications*, DOI: 10.1007/s11042-023-15600-w, 2023.
- [6] Y. Fouzar, A. Lakhssassi and M. Ramakrishna, "A Novel Hybrid Multikey Cryptography Technique for Video Communication," *IEEE Access*, vol. 11, pp. 15693–15700, 2023.
- [7] S. Cheikh Youssef et al., "Video Labelling Robot-assisted Radical Prostatectomy and the Role of Artificial Intelligence (AI): Training a Novice," *Journal of Robotic Surgery*, vol. 17, no. 2, pp. 695–701, DOI:10.1007/s11701-022-01465-y, 2022.
- [8] Z. Zhou and Y. Liu, "Political Teaching Application in High Vocational Care Courses Based on Machine Learning Systems," *Soft Computing*, vol. 27, no. 11, pp. 7657–7666, 2023.
- [9] A. Ayub Khan et al., "BIOmT: A State-of-the-art Consortium Serverless Network Architecture for Healthcare System Using Blockchain Smart Contracts," *IEEE Access*, vol. 10, pp. 78887–78898, DOI:10.1109/access.2022.3194195, 2022.
- [10] A. Ayub Khan et al., "Internet of Things (IoT) Security with Blockchain Technology: A State-of-the-art Review," *IEEE Access*, vol. 10, pp. 122679–122695, DOI:10.1109/access.2022.3223370, 2022.
- [11] J. Yu, H. Chen, K. Wu, T. Zhou, Z. Cai and F. Liu, "Centipede: Leveraging the Distributed Camera Crowd for Cooperative Video Data Storage," *IEEE Internet of Things Jour.*, vol. 8, no. 22, pp. 16498–16509, 2021.
- [12] Y. Wu, Y. Zhang, C. Wang, C. Li and C. Xing, "HuaVideo: Towards a Secure, Scalable and Compatible HTML5 Video Providing System," *Proc. of the 2014 11th IEEE Web Information System and Application Conf.*, DOI:10.1109/wisa.2014.23, Tianjin, China, 2014.
- [13] K. K. Mar, C. Y. Law and V. Chin, "Secure Personal Cloud Storage," *Proc. of the 2015 10th IEEE Int. Conf. for Internet Technology and Secured Transactions (ICITST)*, pp. 108-113, DOI:10.1109/icitst.2015.7412068, London, UK, 2015.
- [14] R. Ma, Q. Wang, Y. Zhang, J. Chen and J. Zhou, "A Dynamic Video Switching Mechanism for Distributed Crane Control System," *Proc. of the 2021 IEEE Intl. Conf. on Dependable, Autonomic and Secure Computing, Intl. Conf. on Pervasive Intelligence and Computing, Intl. Conf. on Cloud and Big Data Computing, Intl. Conf. on Cyber Science and Technology Congress (DASC/PiCom/CBDCOM/CyberSciTech)*, AB, Canada, DOI:10.1109/dasc-picom-cbdcom-cyberscitech52372.2021.00050, 2021.
- [15] D. Ambani and K. Atkotiya, "Secure Data Contribution and Retrieval in Social Networks Using Effective Privacy Preserving Data Mining Techniques," *Proc. of the 2021 IEEE Int. Conf. on Computing, Communication and Green Engineering (CCGE)*, pp. 1-5, DOI:10.1109/ccge50943.2021.9776454, Pune, India, 2021.
- [16] L. Besancon, C. F. Silva and P. Ghodous, "Towards Blockchain Interoperability: Improving Video Games Data Exchange," *Proc. of the 2019 IEEE Int. Conf. on Blockchain and Cryptocurrency (ICBC)*, pp. 81-85, DOI:10.1109/bloc.2019.8751347, Seoul, S. Korea, 2019.
- [17] K. Balint, "Modern, Decentralized Blockchain-based Solutions for Saving Video Footage," *Proc. of the 2020 IEEE 18th International Symposium on Intelligent Systems and Informatics (SISY)*, pp. 11-14, DOI:10.1109/sisy50555.2020.9217100, Subotica, Serbia, 2020.
- [18] L. Du and Y. Li, "Privacy Preserving for Human Object in Video Surveillance *via* Visual Cryptography," *Proc. of the 2014 IEEE Int. Conf. on Security, Pattern Analysis and Cybernetics (SPAC)*, pp. 80-85, DOI:10.1109/spac.2014.6982661, Wuhan, China, 2014.
- [19] V. Shukla and A. Mishra, "A New Sequential Coding Method for Secure Data Communication," *Proc. of the 2020 IEEE Int. Conf. on Computing, Power and Communication Technologies (GUCON)*, pp. 529-533, DOI:10.1109/gucon48875.2020.9231252, Greater Noida, India, 2020.
- [20] C. Jacob and V. R. Rekha, "Secured and Reliable File Sharing System with De-duplication Using Erasure Correction Code," *Proc. of 2017 IEEE Int. Conf. on Networks & Advances in Computational Technologies (NetACT)*, pp. 221-228, DOI:10.1109/netact.2017.8076770, Thiruvananthapuram, India 2017.
- [21] A. Fitwi, Y. Chen and S. Zhu, "A Lightweight Blockchain-based Privacy Protection for Smart Surveillance at the Edge," *Proc. of the 2019 IEEE Int. Conf. on Blockchain (Blockchain)*, pp. 552-555, DOI:10.1109/blockchain.2019.00080, Atlanta, USA, 2019.
- [22] Y.-T. Chen, C.-C. Yen, Y.-T. Lin and J.-S. Wang, "Cooperative Caching Plan of Popular Videos for Mobile Users by Grouping Preferences," *Proc. of the 2018 IEEE 16th Intl. Conf. on Dependable, Autonomic and Secure Computing, 16th Intl. Conf. on Pervasive Intelligence and Computing, 4th Intl. Conf. on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech)*, pp. 762-769, DOI: 10.1109/dasc/picom/datacom/cyberscitech.2018.00132, Athens, Greece, 2018.
- [23] C. W. Ling, A. Datta and J. Xu, "A Case for Distributed Multilevel Storage Infrastructure for Visual Surveillance in Intelligent Transportation Networks," *IEEE Internet Computing*, vol. 22, no. 1, pp. 42–51, DOI:10.1109/mic.2017.265102316, 2018.
- [24] S. Harnal and R. K. Chauhan, "Comparison for Confidential Cryptography in Multimedia Cloud

- Environment," Proc. of the 2018 5th IEEE Int. Conf. on Parallel, Distributed and Grid Computing (PDGC), pp. 148-152, DOI:10.1109/pdgc.2018.8745909, Solan, India, 2018.
- [25] A. A. Khan et al., "A Blockchain Security Module for Brain-computer Interface (BCI) with Multimedia Life Cycle Framework (MLCF)," Neuroscience Informatics, vol. 2, no. 1, p. 100030, DOI: 10.1016/j.neuri.2021.100030, 2022.
- [26] M. Yin, M. Wang, Y. Huo and Z. Rao, "Simulation of Solid-liquid Phase Change at Pore Scale Using Lattice Boltzmann Method with Central Moments in Thermal Energy Storage," Journal of Energy Storage, vol. 49, p. 104116, DOI: 10.1016/j.est.2022.104116, 2022.
- [27] N. A. Ugochukwu et al., "An Innovative Blockchain-based Secured Logistics Management Architecture: Utilizing an RSA Asymmetric Encryption Method," Mathematics, vol. 10, no. 24, p. 4670, DOI: 10.3390/math10244670, 2022.
- [28] J. Tang et al., "Sae-PPL: Self-guided Attention Encoder with Prior Knowledge-guided Pseudo Labels for Weakly Supervised Video Anomaly Detection," Journal of Visual Communication and Image Representation, vol. 97, p. 103967, DOI: 10.1016/j.jvcir.2023.103967, 2023.
- [29] S. Wang, E. Zhu, J. Yin and F. Porikli, "Video Anomaly Detection and Localization by Local Motion Based Joint Video Representation and OCELM," Neurocomputing, vol. 277, pp. 161–175, DOI: 10.1016/j.neucom.2016.08.156, 2018.
- [30] S.-A. Jin, T. Kämäräinen, P. Rinke, O. J. Rojas and M. Todorović, "Machine Learning as a Tool to Engineer Microstructures: Morphological Prediction of Tannin-based Colloids Using Bayesian Surrogate Models," MRS Bulletin, vol. 47, no. 1, pp. 29–37, DOI: 10.1557/s43577-021-00183-4, 2022.

ملخص البحث:

تُعالج الأتمتة المتعلقة بالتوزيع والتخزين الآمن لبيانات الفيديو أوجه القصور في العمليات اليدوية. وينسجم التوزيع والتخزين الآمن للبيانات مع الأهمية المتزايدة لحماية بيانات الفيديو. وتواجه الطرق التقليدية في توزيع بيانات الفيديو وتخزينها جملةً من المحددات، منها نقص القابلية للتوسيع، وازدواجية البيانات، ومشاكل الخصوصية، والقضايا المتعلقة بالزمن (سرعة الكتابة والقراءة)، وفعالية التخزين.

تحاول هذه الدراسة معالجة تلك المحددات وتقليل آثارها على أنظمة توزيع بيانات الفيديو وتخزينها من خلال تحقيق التوزيع الأمثل لبيانات الفيديو، مع تحسين الأمور المتعلقة بالتكرار وسماحية الخطأ، وضمان التوافر المستمر للبيانات حتى في حالات الفشل. يُضاف إلى ذلك، التركيز على بساطة النظام وسهولة الوصول إلى البيانات، خاصة عند التعامل مع مجموعات بيانات الفيديو الضخمة ومالها من تطبيقات متنوعة بما فيها الشبكات الاجتماعية وألعاب الفيديو وشبكات النقل.

وقد تمت مقارنة النظام المقترح مع عدد من الأنظمة الأخرى للتوزيع والتخزين الآمن لبيانات الفيديو، واتضح تفوق النظام المقترح على الكثير من الأنظمة التقليدية؛ فقد بلغ معدل زمن إزالة التشفير للنظام المقترح (0.208) ثانية، ومعدل زمن التشفير (0.182) ثانية، ومعدل زمن التوزيع (1.620) ثانية، وكانت الأقل مقارنةً بأزمنة عمل غيرها من الأنظمة.

ARABIC SOFT SPELLING CORRECTION WITH T5

Mohammed Al-Qaraghuli and Ola Arif Jaafar

(Received: 12-Nov.-2023, Revised: 3-Jan.-2024, Accepted: 20-Jan.-2024)

ABSTRACT

Spelling correction is considered a challenging task for resource-scarce languages. The Arabic language is one of these resource-scarce languages, which suffers from the absence of a large spelling correction dataset, thus datasets injected with artificial errors are used to overcome this problem. In this paper, we trained the Text-to-Text Transfer Transformer (T5) model using artificial errors to correct Arabic soft spelling mistakes. Our T5 model can correct 97.8% of the artificial errors that were injected into the test set. Additionally, our T5 model achieves a character error rate (CER) of 0.77% on a set that contains real soft spelling mistakes. We achieved these results using a 4-layer T5 model trained with a 90% error injection rate, with a maximum sequence length of 300 characters.

KEYWORDS

Arabic spelling correction, Transformers, Text-to-text transfer transformer, T5, Natural-language processing.

1. INTRODUCTION

The Arabic language suffers from various types of spelling mistakes. Most of these mistakes occur due to the complex rules and the various shapes of certain letters. Soft mistakes are one of the most common spelling mistakes that deal with confusion among different shapes of certain letters. For example, the letter *alef* (ا) has two shapes (ا) and (ى) at the end of a word, like the word (عصا) which is a noun that means a stick and the word (عصى) which is a verb that means disobey. Mixing the letter *alef* in this type of words can change the meaning or make the sentence ambiguous. *Al-hamza* (ء) can be written in different shapes (ء, ا, ا, ا, ا, ا, ا) depending on the rules, like the word (قراءة) which is often incorrectly written as (قراءة). Letter *teh* (ت) can be written as (ت) or (ة) at the end of a word, like the word (ذروة); when it's indefinite and added to a definite word such as (ذروة العمل), the sound of letter *teh* indicates that it's written like this (ذروت العمل). Therefore, people tend to mix between the two shapes of the letter *teh*. Also, the letter *teh marbuta* (ة) is incorrectly written as the letter *heh* (ه) at the end of a word, like the word (ساعة) that's written often like (ساعه). Additionally, the insertion and omission of the letter *alef* (ا) after *waw aljamaea* (واو الجماعة) is also a common soft spelling mistake.

Table 1. Targeted characters and their romanization.

Characters	Romanization
ء	<i>Al-hamza</i>
ا	<i>alef</i>
ت	<i>teh</i>
ه	<i>heh</i>
و	<i>waw</i>

Al-Ameri (2015) conducted a study that shows which errors are the most frequent among a group of students in a teaching institute. The number of students that participated in the study was 100 students (40 males and 60 females). In Table 2, we show the frequent errors that occurred in the study. We noticed that errors related to *Al-hamza* are the most frequent. Additionally, the errors that are related to the shape of *alef* occurred with a high percentage among the participants [1].

Awad (2012) performed a study that shows the common spelling mistakes among 130 middle-school students. As shown in Table 3, most of these errors are related to *Al-hamza*. Additionally, the insertion and omission of *alef* after *waw aljamaea* is the second most common error in the study [2].

Table 2. The frequent spelling mistakes as reported in Al-Ameri study.

Index	Spelling Error Type	Percentage
1	Writing <i>hamza mutwsita</i> on an <i>alef</i>	73%
2	Writing <i>alef maqswra</i> instead of <i>alef mamdwda</i>	71%
3	Writing <i>alef mamdwda</i> instead of <i>alef maqswra</i>	70%
4	Omitting <i>alef</i> following a <i>waw</i> at the end of some verb forms	67%
5	Writing <i>teh</i> instead of <i>teh marbuta</i>	67%
6	Writing <i>hamza mutwsita</i> on <i>waw</i>	64%
7	Writing <i>hamza</i> at the end of the word on <i>alef</i>	51%
8	Writing <i>hamza</i> on the line at the end of the word	47%
9	Writing <i>hamza</i> at the end of the word on <i>yeh</i>	47%
10	Dropping <i>lam</i> before the "solar letter"	38%
11	Writing <i>teh marbuta</i> instead of <i>teh</i>	37%
12	Writing <i>hamza mutwsita</i> on <i>yeh</i>	30%
13	Writing <i>hamza alqate</i> instead of <i>hamza Alwasl</i>	28%
14	Inserting <i>alef</i> after <i>waw</i> at the end of a word	25%

Table 3. The frequent spelling mistakes in Awad study.

Index	Spelling Error Type	Percentage
1	Confusing between <i>dād</i> and <i>dha</i>	60%
2	<i>alef</i> after <i>waw aljamaea</i>	59%
3	<i>Al-hamza Al-mutwsita</i>	58%
4	<i>Al-hamza Al-mutatarifa</i>	58%
5	Confusing between <i>teh</i> and <i>teh marbuta</i>	57%
6	Letters pronounced but not written	57%
7	<i>heh marbuta</i>	56%
8	Confusing between <i>hamza Al-wasl</i> and <i>hamza Al-qate</i>	54%
9	Letters written but not pronounced	52%
10	Confusing between solar and lunar <i>lam</i>	42%
11	<i>Al-tanwin</i>	44%
12	<i>Al-hamza Al-awilia</i>	23%

In Table 4, we show some of the soft spelling mistakes that occurred in Arabic company reviews written by customers [3]. We notice that most of these errors are related to *Al-hamza*, since people tend to forget to add it or use the incorrect shape of it. Additionally, we notice another common mistake which is the incorrect use of the letter (هـ) instead of the letter (هـ).

As illustrated above, soft spelling mistakes are common among various Arabic speakers regardless of age and education level and to continue the efforts to provide modern tools to help Arabic speakers and Arabic learners produce error-free text, we present in this paper a Text-to-Text Transfer Transformer (T5) model that automatically corrects Arabic soft spelling mistakes at a character level.

Table 4. Samples of soft spelling mistakes in arabic company reviews dataset.

Error type	Sentence
1. Incorrect shape of <i>Al-hamza</i> . The correct form is (أسعارهم)	اسعارهم ¹ اغلا ² من المحلات بكثير
2. Incorrect shape of <i>Al-hamza</i> and incorrect shape of <i>alef</i> at the end of the word. The correct form is (أعلى)	و بحظولك توصيل مجاني حكي فاضي التطبيق لا انصح ³ به
3. Incorrect shape of <i>Al-hamza</i> . The correct form is (أنصح)	
4. Incorrect shape of <i>Al-hamza</i> . The correct form is (أكثر)	
5. Incorrect shape of <i>Al-hamza</i> . The correct form is (الإجراءات)	برنامج اكثر ⁴ من رائع للتقسيت
6. Using the letter (هـ) instead of the letter (هـ). The correct form is (المعاملة)	سهولة الاجراءات ⁵ وحسن المعاملة ⁶ والدقة ⁷ واحترام العميل
7. Using the letter (هـ) instead of the letter (هـ). The correct form is (الدقة)	
8. Incorrect shape of <i>Al-hamza</i> and incorrect insertion of (ل) after the letter (و) at the end of the word. The correct form is (أرجو)	ارجوا ⁸ تشغيل خط من الحرم اليوناني الجامعة الأمريكية بوسط البلد الى ⁹ المقطم شارع 9
9. Incorrect shape of <i>Al-hamza</i> . The correct form is (الى)	

Previous works that handle Arabic soft spelling mistakes used both BiLSTM and the original transformer and the T5 model has overtaken the original transformer as the way-to-go encoder-decoder model. T5 has a simpler architecture than the original transformer, yet achieved better results compared to the other transformer models in various tasks [4][5][6]. In this paper, T5 shows a CER reduction of 10.4% over the previous works that used the original transformer and handled the same type of errors.

The rest of the paper is structured as follows: In Section 2, we will review the most recent works that are related to text correction. In Section 3, we give an overview of the T5 model. In Section 4, we show our work methodology. In Section 5, we show our results and discuss them. In Section 6, we state the limitations of our work. Finally, in Section 7, we give the conclusion and our future work ideas.

2. LITERATURE REVIEW

Recent works in both Arabic and foreign languages started to use transformer neural networks in text-correction tasks, whether being spelling correction, grammatical correction or post-ASR correction. In this section, we will review the most recent works that dealt with text correction for Arabic and foreign languages.

2.1 Arabic Text Correction

Al-Oyaynaa and Kotb introduced a detection system for Arabic grammatical errors using the Arabic version of the BERT model (AraBERT) and the Multilingual BERT (M-BERT). They tackled the problem at the token level and sentence level and their best results were achieved using the Arabert model with an F_1 score of 87% at the token level and an F_1 score of 98% at the sentence level [7]. Abandah *et al.* proposed an error-injection approach called stochastic error Injection to insert artificial errors into a correct text, thus providing the model with enough data to train. They used this approach to train a BiLSTM model to correct Arabic soft spelling mistakes. They achieved a CER of 1.28% on a set with real soft errors called Test200, using a 2-layer BiLSTM model with a 40% error injection rate [8]. Similar to [8], Al-Qaraghuli *et al.* introduced a transformer-based model trained from scratch to correct Arabic soft spelling mistakes. They used the original transformer architecture and trained it on a large text from Wikipedia that was injected with artificial errors using stochastic error Injection. They achieved a CER of 0.86% on the Test200 set, using a 4-layer transformer model with a 90% error injection rate [9]. Madi and Al-Khalifa introduced three models to detect and correct a variety of Arabic errors, such as syntax errors, semantic errors and spelling errors. They achieved an $F_{0.5}$ score of 81.55% using the BiLSTM model [10].

2.2 Foreign Text Correction

Wei *et al.* proposed a detection and correction system for Chinese spelling mistakes. They built the detection part of the system based on the ELECTRA model. As for the correction part, they implemented three models based on BERT. They evaluated their system on three datasets achieving an average of 5.8% F_1 improvement over previous works [11]. Stankevičius *et al.* proposed a multilingual model based on ByT5 and statistical Unigram to correct typographical errors and restore diacritics. They achieved a 94.6% average accuracy for 13 languages [12]. Neto *et al.* introduced a spelling-correction model to recognize a handwritten text in English, French and Latin. They implemented an encoder-decoder model with the Luong attention mechanism. They achieved a CER of 3.2% and a WER of 7.7% [13].

3. TEXT-TO-TEXT TRANSFER TRANSFORMER

Text-to-Text Transfer Transformer (T5) is an encoder-decoder transformer model proposed by Raffel *et al.* [14]. The main idea of the T5 model is to leverage transfer learning to produce a unified framework for multiple NLP tasks. T5 is built with a similar architecture to the original transformer that was proposed by Vaswani *et al.* [15]. The major difference between T5 and the original transformer is the positional encoding method, where T5 uses a simplified relative positional encoding, while the original transformer uses sinusoidal positional encoding. The other difference is that T5 removes the additive bias in the normalization layer.

T5 was pre-trained on the Colossal Clean Crawled Corpus (C4) dataset [14] and then finetuned on multiple NLP tasks. In the finetuning stage, T5 uses a prefix to identify the task that is required to work on it.

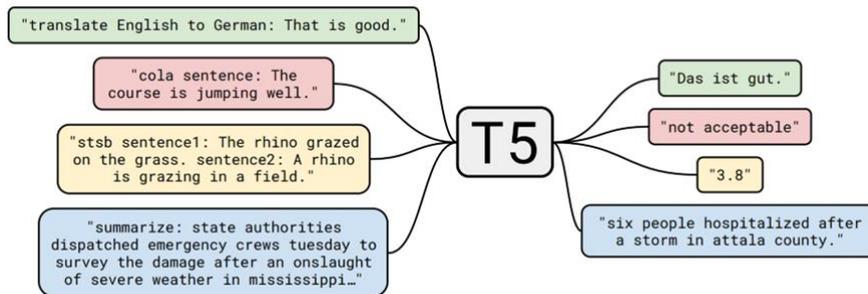


Figure 1. NLP tasks and their prefixes in T5 model.

As shown in Figure 1, we can notice the prefixes such as “translate” which tells the model to translate text. T5 can translate English text into three languages; namely, German, French and Romanian. The second prefix is “summarize” which tells the model to summarize the text. The other prefixes are named based on the task’s dataset such as “cola sentence” which answers whether the sentence is grammatically acceptable or not. As for “stsb sentence 1 sentence 2”, it’s a prefix used for sentence-similarity tasks. In this task, the model outputs a number as a string to measure whether sentence 1 is similar to sentence 2 or not.

Additionally, the authors of T5 introduced an improved version of T5 called T5 version 1.1. This version uses the GEGLU activation function instead of ReLU. Also, the embedding layer doesn’t share parameters with the classifier layer.

It is worth mentioning that T5 V1.1 was only pre-trained on the C4 dataset, unlike the regular T5 which was pre-trained and finetuned on the previously mentioned tasks.

4. METHODOLOGY

In this section, we show how we trained our T5 model and the datasets we used in the training and evaluation. For training, we injected the Wiki-40B set with artificial errors and trained the T5 model on the set. Then, we finetuned our model by changing the number of layers, the error injection rate and the maximum sequence length. For evaluation, we used two sets; the Wiki-40B test set and the Test200 set. We used two evaluation metrics; BLEU Score and CER. We calculated the BLEU Score on the Wiki-40B test set and CER on Test200. Figure 2 summarizes our work methodology.

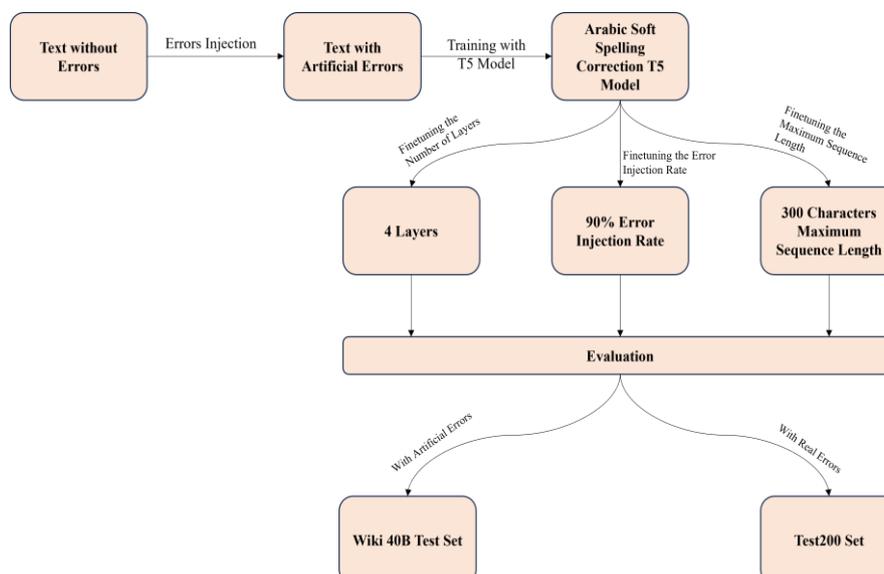


Figure 2. The methodology of our work.

4.1 Model Building and Training

We built our T5 model using three configurations, as described in Table 5. Configuration 3 has the same parameters as the small T5 in [14].

Table 5. T5 configurations.

Parameter	Conf. 1	Conf. 2	Conf. 3
Model Dimension	128	512	512
Feed Forward Dimension	512	1024	2048
Number of Heads	8	8	8
Encoder Layers	2	4	6
Decoder Layers	2	4	6
Dropout Rate	0.1	0.1	0.1
Batch Size	128	128	64
Model Parameters	2m	21m	44m

All three configurations were trained with Kaggle TPUs. The specifications of the Kaggle platform are reported in Table 6. In the training stage, we used Adam optimizer with a learning rate = $1e-4$, sparse categorical crossentropy as a loss function and used Accuracy as a training and validation metric.

In Table 7, we report our training and validation stats for the three configurations.

Table 6. Kaggle-platform specifications.

Aspect	Specification
CPU	Intel(R) Xeon(R) CPU @ 2.20GHz
TPU	v3-8
GPU	Nvidia P100 @ 1.32GHz, 16 GB
Memory	30 GB
Libraries	Python 3.10.12, TensorFlow 2.13.0

Table 7. The training and validation stats for the three configurations.

Parameter	Conf. 1	Conf. 2	Conf. 3
Number of Epochs	104	53	36
Training Time (Hours)	21.66	27.51	33.24
Best Accuracy on the Train Set	99.60%	99.73%	99.75%
Best Accuracy on the Validation Set	99.66%	99.71%	99.71%
Best Loss on the Train Set	1.29%	0.85%	0.79%
Best Loss on the Validation Set	1.14%	0.99%	1.02%

4.2 Datasets

In this work, we trained our T5 model using a dataset named Wiki-40B; for evaluation, however, we used two sets; the first one is the Wiki-40B test set (only 2k sequences were selected, because text generation will take a long time and requires using more resources) and the second one is Test200. The details of each set and preparation steps are described in the following sub-sections.

4.2.1 Wiki-40B

Wiki-40B is a multilingual dataset with 40 billion characters [16]. Wiki-40B Arabic version has 245,354 articles that were split into three sub-sets: train, validation and test.

In Table 8, we show the occurrences of the target characters in the Wiki-40B set. Character (l) is the most occurring character with 14.27% while the character (⤵) is the least occurring character with 0.04%.

Table 8. Targeted character occurrences in Wiki-40B set.

Character(s)	Wiki-40B Set
ء	0.04%
أ	1.94%
ئ	0.40%
إ	0.75%
ؤ	0.09%
ا	14.27%
آ	0.10%
ة	3.41%
ى	0.81%
و	0.46%
ت	1.16%
وا	0.07%
ه	0.70%
اء	0.27%
Total	24.47%

We prepared the Wiki-40B set for our model as follows:

1. We removed punctuation marks, numbers and English letters, thus only Arabic letters remained in the set.
2. We wrapped all sequences that are longer than 300 characters, which is the selected maximum sequence length, to optimize model training.
3. Lastly, we injected artificial errors into the dataset using the stochastic error-injection approach that was proposed in [8].

Table 9. Wiki-40B transformation after being processed.

Original Text	Transformed Text
<p>_START_ARTICLE_ احتلال الفضاء الإلكتروني</p> <p>_START_PARAGRAPH_ يُقصد بمصطلح احتلال الفضاء الإلكتروني (الذي يُعرف كذلك باسم احتلال النطاق الإلكتروني)، وفقاً للقانون الفيدرالي الأمريكي الذي يحمل اسم قانون حماية المستهلك ومكافحة احتلال الفضاء الإلكتروني، تسجيل اسم نطاق أو الاتجار فيه أو استخدامه بنية غير حسنة للربح من شهرة علامة تجارية تخص شخصاً آخر، حيث يعرض محتل الفضاء الإلكتروني بيع النطاق للشخص أو الشركة التي تمتلك العلامة التجارية الواردة ضمن الاسم بسعر مبالغ فيه</p> <p>وهذا المصطلح مشتق من لفظ الاحتلال الذي يعني وضع اليد على أرض أو مبنى شاغر أو مهجور لا يمتلكه المحتل ولم يتم باستجاره ولم يمنح الإذن باستخدامه. غير أن احتلال الفضاء الإلكتروني يختلف عن ذلك إلى حد ما؛ إذ يتم أحياناً دفع مقابل أسماء النطاق التي يتم "احتلالها" عن طريق التسجيل التي يقوم بها محتلو الفضاء الإلكتروني وعادة ما يطلب محتلو الفضاء الإلكتروني أسعاراً أعلى من تلك التي اشترى بها. كما يضع بعض محتلو الفضاء الإلكتروني تعليقات تنتقص من شأن الشخص أو الشركة التي يعبر عنها النطاق، سعياً لتشجيع الضحية على النطاق من المحتلين، في حين يقوم آخرون بإرسال روابط مدفوعة عبر جوجل وبهاو وأسك دوت كوم وغيرها من شبكات الإعلان مدفوعة الأجر، إلى الموقع الفعلي الذي كان المستخدم يريده، ومن ثم يقننون احتلالهم.</p>	<p>احتلال الفضاء الإلكتروني</p> <p>يقصد بمصطلح احتلال الفضاء الإلكتروني الذي يعرف كذلك باسم احتلال النطاق الإلكتروني وفقاً للقانون الفيدرالي الأمريكي الذي يحمل اسم قانون حماية المستهلك ومكافحة احتلال الفضاء الإلكتروني تسجيل اسم نطاق أو الاتجار فيه أو استخدامه بنية غير حسنة للربح من شهرة علامة تجارية تخص شخصاً آخر، حيث يعرض محتل الفضاء الإلكتروني بيع النطاق للشخص أو الشركة التي تمتلك العلامة التجارية الواردة ضمن الاسم بسعر مبالغ فيه</p> <p>وهذا المصطلح مشتق من لفظ الاحتلال الذي يعني وضع اليد على أرض أو مبنى شاغر أو مهجور لا يمتلكه المحتل ولم يتم باستجاره ولم يمنح الإذن باستخدامه. غير أن احتلال الفضاء الإلكتروني يختلف عن ذلك إلى حد ما؛ إذ يتم أحياناً دفع مقابل أسماء النطاق التي يتم احتلالها عن طريق عملية التسجيل التي يقوم بها محتلو الفضاء الإلكتروني وعادة ما يطلب محتلو الفضاء الإلكتروني أسعاراً أعلى من تلك التي اشترى بها.</p> <p>كما يضع بعض محتلي الفضاء الإلكتروني تعليقات تنتقص من شأن الشخص أو الشركة التي يعبر عنها النطاق سعياً لتشجيع الضحية على شراء النطاق من المحتلين، في حين يقوم آخرون بإرسال روابط مدفوعة عبر جوجل وبهاو وأسك دوت كوم وغيرها من شبكات الإعلان مدفوعة الأجر، إلى الموقع الفعلي الذي كان المستخدم يريده ومن ثم يقننون احتلالهم.</p>

In Table 9, we show the transformation of the Wiki-40B dataset after being processed. We transformed the set from long articles to short and understandable sentences. In Table 10, we show

Wiki-40B and Test200 characteristics. The dataset contains 4m sequences now which is considered a suitable size to train a transformer model.

4.2.2 Test200

We used Test200 as our evaluation set, because it contains real soft mistakes that were collected and corrected manually. The set has two hundred sequences with a total number of errors of 1,306 and with an average of 6.5 errors per sequence [8].

Table 10. Wiki-40B and Test200 characteristics.

Criterion	Wiki-40B Set	Test200 Set
Sequence count	4m	200
Word count	73.5m	2,443
Arabic letter count	354m	24,002
Words per sequence	18.3	12.2
Letters per word	4.8	9.8

4.3 Evaluation Metrics

We used two evaluation metrics; the first one is character error rate (CER) to show whether the model can correct soft spelling errors, and the lower the better the CER value.

The second metric is bilingual evaluation understudy (BLEU) to measure whether the model can produce a structured text similar to the target text [17]. The higher the better the BLEU value.

5. RESULTS AND DISCUSSION

In this section, we show the four experiments that we conducted and the results that we obtained. We finetuned our model by changing the number of layers, using three error-injection rates and using two different maximum sequence lengths.

5.1 Model Size

In Figure 3, we show the BLEU score of the three configurations on the Wiki-40B test set. All the results are high, which indicates that all three configurations can output understandable Arabic text. We notice that configurations 2 and 3 have higher results than configuration 1, which is because the more the number of layers increased the better the model can capture the characteristics of the language and therefore, it can produce structured and understandable text. Additionally, having a high BLEU score is related to the number of targeted letters in the set. We only targeted certain letters; so, the rest of the letters should remain the same without any change. After confirming that all configurations can understand the Arabic language, now we look at CER on the Test200 set to examine which configuration is performing better.

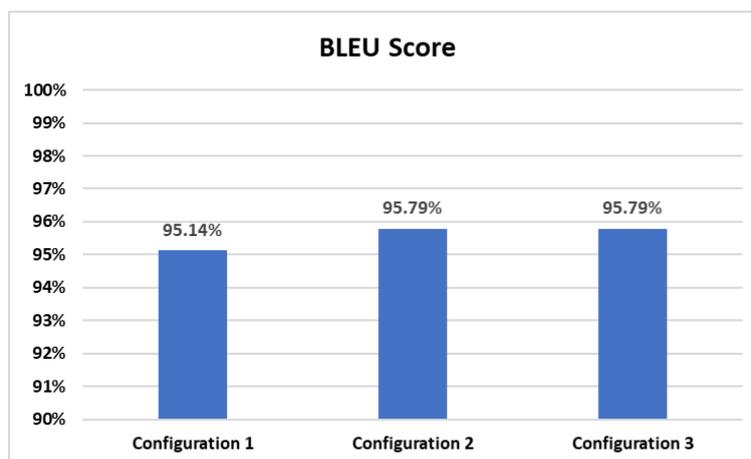


Figure 3. BLEU scores on Wiki-40B test set for the three selected configurations.

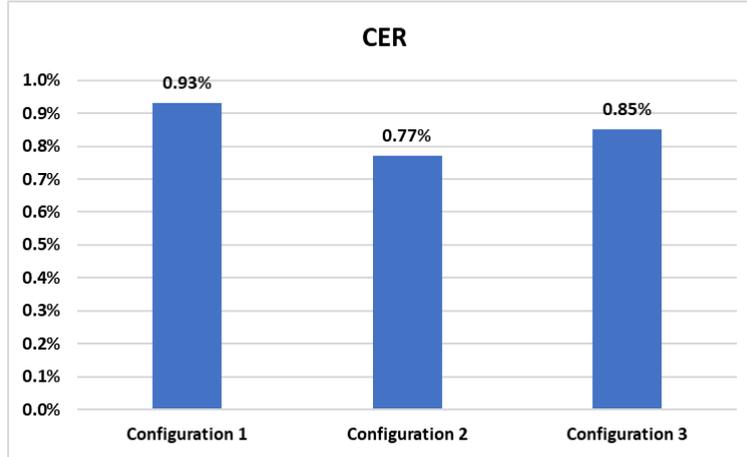


Figure 4. CER results on Test200 set for the three selected configurations.

We can observe in Figure 4 that configuration 2 achieved the lowest CER of 0.77% compared to 0.85% of Configuration 3, while configuration 1 obtained a CER of 0.93%, which is higher compared to both configurations 2 and 3. In contrast to BLEU, CER measures the model's ability to correct the errors in the set; so, the optimal value does not only depend on the number of layers. The size of the training set also plays a major factor in this process. The size of Wiki-40B is suitable for a 4-layer transformer model, as this work shows and the work in [9]. Additionally, if the size of the set is small, transformer models may not be able to outperform BiLSTM models as shown in [9]. Based on these results, we selected configuration 2 to conduct the rest of our experiments.

5.2 Error-injection Rate

In this experiment, we used three error-injection rates; 40%, 60% and 90%. In Figure 5, we see that the 90% rate is indeed the best rate for our T5 model achieving a CER of 0.77%. This result aligns with the one reported in [9] that suggests that the 90% rate is the one suitable for a transformer model, because the high injection rate gives the model more errors to correct and the attention mechanism benefits from this, in contrast to BiLSTM models that required lower injection rates, such as 40% as reported in [8].

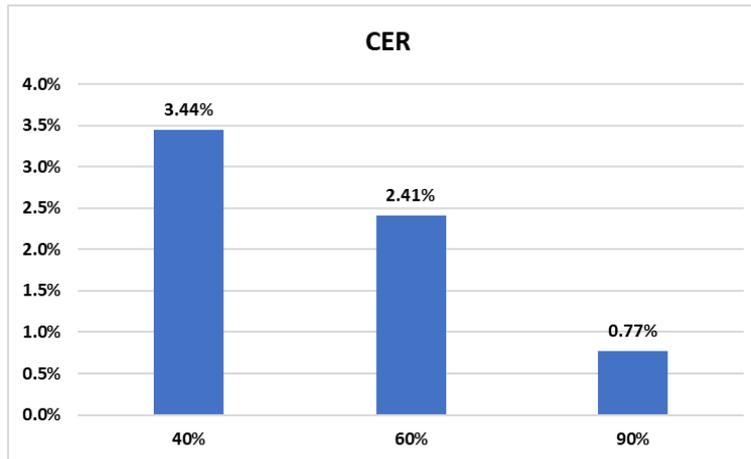


Figure 5. CER results on Test200 set using three different error-injection rates.

5.3 Maximum Sequence Length

The third experiment is related to the maximum sequence length. We tested whether increasing the maximum sequence length can improve the performance of the model. We increased the length to 500 characters; the model obtained a CER of 0.78%. The result is not improved compared to the 0.77% that was obtained using 300 characters as maximum sequence length. We also observed that training time has increased significantly when we increased the maximum sequence length, as shown in Figure

6. The model that was trained with a max. length of 300 characters took 27.51 hours, while the one trained with a max. length of 500 characters took 59.36 hours.

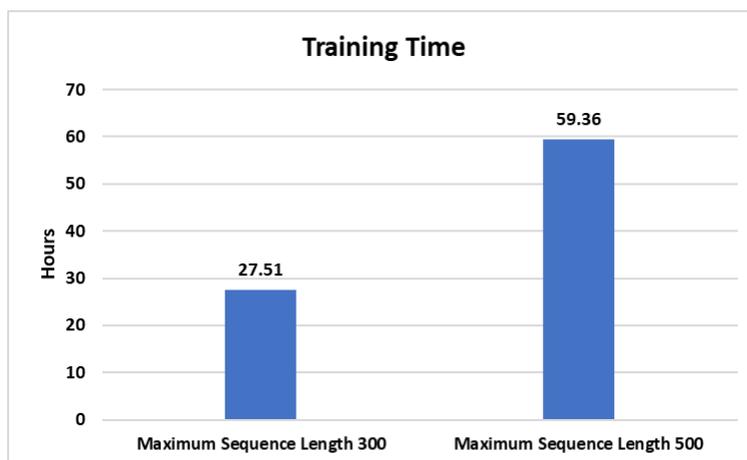


Figure 6. T5 training time in hours using 300 and 500 maximum sequence length.

5.4 T5 Version 1.1

In the final experiment, we used T5 V1.1 to check whether it could lower the results that we obtained so far. As shown in Figure 7, the results did not improve compared to the results of the regular T5.

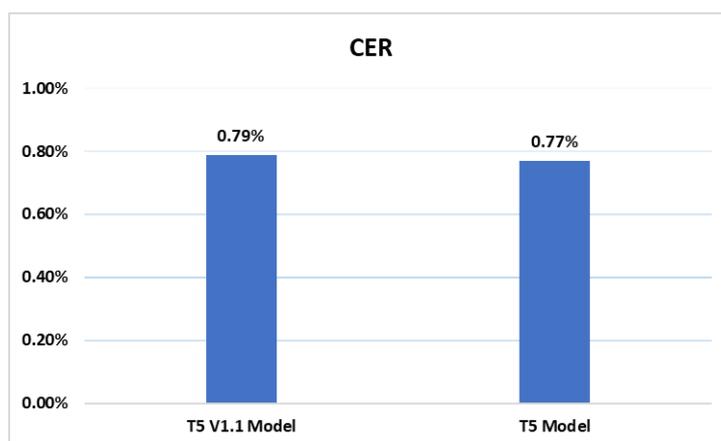


Figure 7. T5 and T5 V1.1 CER results on Test200 set.

In summary, we conducted four experiments and obtained the lowest CER of 0.77% using four layers. T5 model was trained with a 90% error-injection rate and a maximum sequence length of 300 characters.

5.5 Confusion Matrix

In Figure 8, we show the confusion matrix of our best model on the Wiki-40B test set. The set was injected with 32134 artificial errors and our model was able to correct 97.8% of these errors. We can observe that the letter (l) is the one with the most confusion. It was falsely predicted as the letter (ل) 200 times and as the letter (ل) 82 times.

In Table 11, we show the ability of our model to correct real soft errors that were previously explained in Table 4. We can notice that the T5 model can correct the four types of errors that we targeted, which can be shown in the following words: The word (اغلا) contains two types of errors; the first one is *Al-hamza* shape at the beginning of the word and the second one is the *alef* shape at the end of the word. The model was able to correct these errors and restore the right shape of the word (أغلى). The third type of error that we targeted is the shape of *waw* at the end of a word; the word (ارجوا) contains this type of error. The model was able to correct the shape of *waw* at the end of the word (أرجو) along with *Al-hamza* shape at the beginning of the word (أرجو). The last type of error that we targeted is the

shape of *teh* and *heh* at the end of a word. This error appears in two words (المعامله) and (الدقه) and both words were successfully corrected; (المعاملة) and (الدقة).

		Predicted Letter													
		وا	ه	اء	و	ت	ء	آ	أ	ؤ	إ	ئ	ا	ة	ى
Actual Letter	وا	111	0	0	1	0	0	0	0	0	0	0	0	0	0
	ه	0	1146	0	0	15	0	0	0	0	0	0	0	63	0
	اء	0	0	486	0	0	0	0	0	0	0	0	0	0	0
	و	1	0	0	855	0	0	0	0	0	0	0	0	0	0
	ت	0	10	0	0	2325	0	0	0	0	0	0	0	14	0
	ء	0	0	0	0	0	60	0	0	0	0	0	0	0	0
	آ	0	0	0	0	0	0	160	11	0	0	0	4	0	0
	أ	0	0	0	0	0	0	13	3588	1	46	3	40	0	0
	ؤ	0	0	0	0	0	0	0	3	167	0	2	1	0	0
	إ	0	0	0	0	0	0	3	37	1	1323	0	73	0	0
	ئ	0	0	1	0	0	1	0	3	6	0	745	1	0	0
	ا	0	0	0	0	0	0	6	200	5	82	1	26466	0	10
	ة	0	27	0	0	10	0	0	0	0	0	0	0	6473	0
	ى	0	0	0	0	0	0	0	0	0	0	0	12	0	1520

Figure 8. T5 confusion matrix on Wiki-40B test set.

Table 11. Model ability to correct soft spelling mistakes.

Input Sentence	Predicted Sentence
اسعارهم اغلا من المحلات بكثير و بحطولك توصيل مجاني حكي فاضي التطبيق لا انصح به	أسعارهم أغلى من المحلات بكثير و بحطولك توصيل مجاني حكي فاضي التطبيق لا أنصح به
برنامج أكثر من رائع للتقسيت سهولة الإجراءات وحسن المعامله والدقة واحترام العميل	برنامج أكثر من رائع للتقسيت سهولة الإجراءات وحسن المعاملة والدقة واحترام العميل
ارجوا تشغيل خط من الحرم اليوناني الجامعة الأمريكية بوسط البلد الى المقطم شارع 9	أرجو تشغيل خط من الحرم اليوناني الجامعة الأمريكية بوسط البلد إلى المقطم شارع 9

5.6 Comparison

We compared our results with those reported in [8] and [9]. These works used the same training approach and the same set of errors which is the soft errors.

In Figure 9, we can observe that our T5 model obtained the lowest CER of 0.77% compared to 0.86% of the Wiki transformer model and 1.28% of the Tashkeela BiLSTM model. The CER result shows that the T5 model outperforms both the original transformer and BiLSTM.

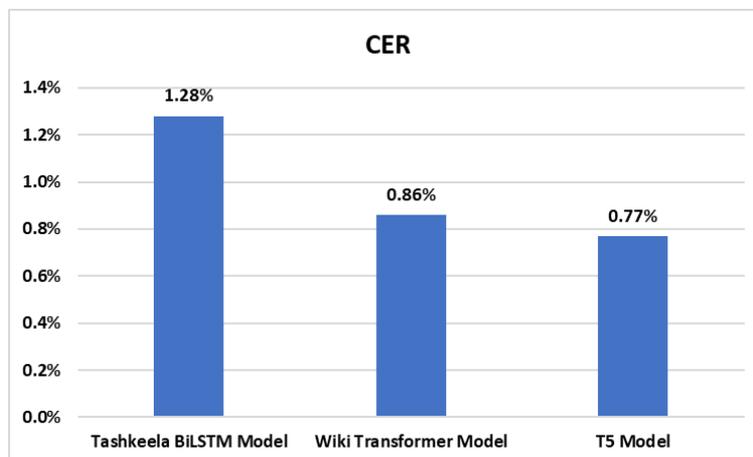


Figure 9. Our T5 CER results on Test200 set compared to the previous results of Wiki transformer model and Tashkeela BiLSTM model.

6. LIMITATIONS

In this work, we showed that the T5 model can correct Arabic soft spelling mistakes better than the original transformer and BiLSTM neural networks. Yet, this work has certain limitations, such as the lack of a large dataset that contains real soft mistakes for both training and evaluation. Additionally, using artificial errors is also limited by the size of both the dataset and the model due to the limited resources that we have. Injecting a very large set and using a large-size model are not possible in our case, requiring dedicated resources.

7. CONCLUSIONS

Nowadays, transformers dominate the natural-language processing field. Many transformer models, such as BERT, BART, T5, GPT2 and GPT3, have become the way-to-go solution for tasks, like machine translation, sentiment analysis, text generation, question answering and spelling correction. In this paper, we implemented a transformer model called T5 to correct Arabic soft spelling mistakes. We corrected four types of soft errors that are related to the shape of a letter; these types are the confusion among *Al-hamza* shapes (أ, إ, ؤ, ء), the confusion among *teh*, *teh marbuta* and *heh* at the end of a word, the confusion between the two shapes of *alef* at the end of a word and the insertion and omission of *alef* after *waw aljamaea*. We achieved optimal results using a four-layer T5 model trained on Wiki-40B set that was injected with artificial errors at a 90% rate. Our model can correct 97.8% of the 32134 artificial errors that were injected into the Wiki-40B test set. We also evaluated our model using real soft errors, where our model achieved a CER of 0.77% on the Test200 set.

For future work, we would like to leverage the unified frame of T5 to create a unified model that can correct more than one type of errors, such as grammatical errors, restore and correct diacritics and other types of spelling mistakes. We are looking to use the prefix aspect of T5 to unify these types of errors and to create a training set or an evaluation set for these errors.

Additionally, we like to investigate the effects of using a bidirectional encoder model, such as BART [18] and BERT [19]. We are also looking to correct spelling mistakes at the word level using the previously mentioned models.

REFERENCES

- [1] A. Al-Ameri, "Common Spelling Mistakes among Students of Teacher Education Institutes," The Islamic College University Journal, vol. 2015, no. 33, pp. 445-474, 2015.
- [2] F. Awad, "Spelling Errors, Their Causes and Methods of Treatment," DIRASAT TARBAWIYA, vol. 5, no. 17, p. 217-250, 2012.
- [3] Kaggle, "Arabic Company Reviews," [Online], Available: <https://www.kaggle.com/datasets/fahdseddik/arabic-company-reviews>.
- [4] T. Adewumi, S. Sabry, N. Abid, F. Liwicki and M. Liwicki, "T5 for Hate Speech, Augmented Data and Ensemble," Sci, vol. 5, no.4, p. 37, DOI: 10.3390/sci5040037, 2023.
- [5] H. Zhuang et al., "RankT5: Fine-tuning T5 for Text Ranking with Ranking Losses," Proc. of the 46th Int. ACM SIGIR Conf. on Research and Development in Information Retrieval, pp. 2308–2313, DOI: 10.1145/3539618.3592047, 2023.
- [6] M. Fu et al., "VulRepair: A T5-based Automated Software Vulnerability Repair," Proc. of the 30th ACM Joint European Software Engineering Conf. and Symposium on the Foundations of Software Engineering, pp. 935-947, DOI: 10.1145/3540250.3549098, 2022.
- [7] S. Aloyayna and Y. Kotb, "Arabic Grammatical Error Detection Using Transformers-based Pre-trained Language Models," ITM Web of Conferences, vol. 56, p. 04009, 2023.
- [8] G. Abandah, A. Suyyagh and M. Z. Khedher, "Correcting Arabic Soft Spelling Mistakes Using BiLSTM-based Machine Learning," Int. J. of Advanced Computer Sci. and Appl., vol. 13, no. 5, 2022.
- [9] M. Al-Qaraghuli, G. Abandah and A. Suyyagh, "Correcting Arabic Soft Spelling Mistakes Using Transformers," Proc. of the 2021 IEEE Jordan Int. Joint Conf. on Electrical Engineering and Information Technology (JEEIT), pp. 146-151, Amman, Jordan, 2021.
- [10] N. Madi and H. Al-Khalifa, "Error Detection for Arabic Text Using Neural Sequence Labeling," Applied Sciences, vol. 10, no. 15, p. 5279, 2020.
- [11] X. Wei, J. Huang, H. Yu and Q. Liu, "PTCSPELL: Pre-trained Corrector Based on Character Shape and Pinyin for Chinese Spelling Correction," Proc. of Findings of the Association for Computational Linguistics: ACL 2023, pp. 6330–6343, Toronto, Canada, 2023.

- [12] L. Stankevičius, M. Lukoševičius, J. Kapočiūtė-Dzikiene, M. Briedienė and T. Krilavičius, "Correcting Diacritics and Typos with a ByT5 Transformer Model," Applied Sciences, vol. 12, no. 5, p. 2636, 2022.
- [13] A. F. de S. Neto, B. L. D. Bezerra and A. H. Toselli, "Towards the Natural Language Processing As Spelling Correction for Offline Handwritten Text Recognition Systems," Applied Sciences, vol. 10, no. 21, p. 7711, 2020.
- [14] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li and P.J. Liu, "Exploring the Limits of Transfer Learning with a Unified Text-to-text Transformer," The Journal of Machine Learning Research, vol. 21, pp. 1-67, 2020.
- [15] A. Vaswani et al., "Attention Is All You Need," Proc. of the 31st Conf. on Neural Information Processing Systems (NIPS2017), pp. 1-11, Long Beach, USA, 2017.
- [16] M. Guo, Z. Dai, D. Vrandečić and R. Al-Rfou, "Wiki-40B: Multilingual Language Model Dataset," Proc. of the 12th Language Resources and Evaluation Conf., pp. 2440–2452, Marseille, France, 2020.
- [17] M. Post, "A Call for Clarity in Reporting BLEU Scores," Proc. of the 3rd Conf. on Machine Translation: Research Papers, pp. 186-191, Brussels, Belgium, 2018.
- [18] M. Lewis et al., "BART: Denoising Sequence-to-sequence Pre-training for Natural Language Generation, Translation and Comprehension," arXiv preprint, arXiv: 1910.13461, 2019.
- [19] J. Devlin, M. Chang, K. Lee and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," arXiv preprint, arXiv: 1810.04805, 2018.

ملخص البحث:

يُعدّ تصحيح أخطاء التهجئة مهمةً تتطوي على تحدياتٍ، وبخاصّة في اللغات نادرة المصادر. واللغة العربية هي إحدى تلك اللغات؛ فهي تعاني من غياب مجموعة بيانات ضخمة لتصحيح أخطاء التهجئة. لذا تُستخدم مجموعات بياناتٍ يتمّ حقنها بأخطاء اصطناعية للتغلب على هذه المعضلة.

في هذه الورقة، فُمنّا بتدريب محوّل التقل من نصّ إلى نصّ (T5) باستخدام أخطاء اصطناعية لتصحيح أخطاء التهجئة "الناعمة" باللغة العربية. وأتضح أنّ نموذج (T5) المستخدم في هذه الدراسة بإمكانه تصحيح ما نسبته 97.8% من الأخطاء الاصطناعية التي جرى حقنها في مجموعة بيانات الاختبار. كذلك فقد حقّق نموذجنا معدّل خطأ أحرف (CER) مقداره 0.77% على مجموعة بياناتٍ تحتوي على أخطاء تهجئة "ناعمة" حقيقية. وقد تمّ الحصول على هذه النتائج باستخدام نموذج (T5) رباعي الطبقات جرى تدريبه بمعدّل حقن أخطاء مقداره 90% وطول تتابع أقصى مقداره 300 حرف.

KOCH FRACTAL OCTAGONAL ANTENNA WITH A COMPACT DESIGN AND DEFECTED GROUND STRUCTURE (DGS) FOR ULTRA-WIDEBAND (UWB) WIRELESS USAGE

Tejaswita Kumari, Abu Nasar Ghazali and Anupama Senapati

(Received: 3-Nov.-2023, Revised: 28-Dec.-2023, Accepted: 21-Jan.-2024)

ABSTRACT

This paper introduces a Koch fractal octagonal antenna designed for the ultra wideband (UWB) frequency range. The utilization of the Koch fractal in the antenna design contributes to size reduction and provides compactness in the antenna for UWB. Additionally, it has been noted that the Koch fractal offers wideband operation. The antenna employs copper as the conductor material and Flame Retardant-4 (FR-4) serves as the substrate. The substrate has a dielectric constant $\epsilon_r = 4.4$, a loss tangent $\tan\delta = 0.02$ and a thickness (h) of 1.6 mm. The overall antenna's dimensions are $34.3 \times 26.5 \times 1.6$ and its electrical dimensions are $0.68\lambda_0 \times 0.53\lambda_0 \times 0.032\lambda_0$. It achieves a maximum gain of 8.94 dBi at a frequency of 12.25 GHz, offering a broad bandwidth ranging from 2 GHz to 12.1 GHz. This antenna exhibits resonance at three distinct frequencies; namely, 3.3 GHz, 6 GHz and 8.6 GHz, making it highly efficient with an overall efficiency of 96.8%. The time-domain characteristic of the antenna is acceptable for UWB, group delay is 1.1 ns, the proposed antenna has a high-fidelity factor of 90.2 and the correlation coefficient is 0.9, which makes the antenna a good candidate for UWB. Due to its performance characteristics, this proposed antenna is well-suited for short-range wireless personal area networks (WPANs), supporting high-data-rate applications, like Bluetooth and wireless USB and Wireless Body Area Network (WBAN) applications. Its proficiency also extends to industrial settings, where it helps with precision in control systems, asset tracking and short-range sensing and radar systems.

KEYWORDS

Koch fractal geometry, UWB, DGS, Group delay, Correlation coefficient, Fidelity factor, Octagonal, WBAN, WPAN.

1. INTRODUCTION

In the modern era, ultra-wideband antennas encounter many challenges, such as miniaturization, cost-effectiveness, small size, mechanical robustness, attaining high-performance characteristics, like good gain and large bandwidth to accommodate wireless security and wireless data. Due to the need for overcoming these issues, developing a UWB antenna for wireless communication is complex in the operational bandwidth of 3.1-10.6 GHz as per FCC [1]-[2].

The abovementioned limitations of UWB antenna, can be resolved by using fractal geometry. Fractal UWB antennas can be leveraged in IoT for data transmission and reception, sensing, location identification and positioning. In UWB antenna design, the main focus is on compactness, miniaturization and wide bandwidth and the design should be direction-independent.

Therefore, the properties of self-similarity and space-filling make fractal geometry a good candidate for UWB antenna design [3], because fractal geometry leads to an enhancement in the antenna's performance. From the previous research done by many researchers [4][5][6], it's known that fractal main sub-classes are mass and boundary. Mass fractal is used for wideband multiband applications. There are so many fractal structures, like Koch, Hilbert, Minkowski, Peano and Sierpinski [4] for reducing the size of the antenna, miniaturization and wideband. Asymmetrical shape and curvature in the geometry of fractals bring the eyes of researchers as well as industries. It is marked that the miniaturization process is easily done with the help of fractal geometry because of the self-similarity characteristics and a feature that fills space [7]-[8]. Fractal geometries resembling Koch snowflake [7][8][9], Minkowski [10]-[11] and Sierpinski triangles [4][7][12], are studied for UWB antenna design. For generation, multiband Koch fractal is used in [16] and in [17], the Koch fractal iteration is used for

the generation of the WLAN frequency. Tapered tree-shaped fractals for wireless applications in [19]-[20] are used for UWB. In [22], fractal ring resonator is used and in [23], fractal geometry is used for compactness for UWB applications. There are different types of UWB antennas reported recently [24] that use a DGS technique for achieving super wide band, [25] uses a transparent antenna implanted on a solar-cell substrate and achieves super wideband. In [26], a stingray-shaped antenna is introduced and the CPW technique is used, where the peak gain of the antenna is 3.5 dBi. There are various antennas reported [27][28][29] that use DGS and CPW techniques to achieve the UWB bandwidth. [30] introduces a metasurface-based MIMO antenna, in this antenna, the isolation is enhanced by using the shorting pin and slot. The diversity gain is 9.99 dB.

It is observed that because of discontinuities and curves existing in the fractal antenna, the electromagnetic radiation is very proficient. Discontinuities and curves provide alteration in the current path, because of which the antenna properties improve, especially radiation [5]. The conception that the Koch fractal geometry would yield a UWB outcome is based on these inherent properties of fractal structures, like multiband properties, miniaturization and space-filling, as well as enhanced bandwidth, improved radiation characteristics and frequency-independent properties researchers and antenna designers have explored various fractal geometries due to their unique characteristics, such as multiband operation, compact size and broadband capabilities, which align well with the requirements of UWB systems. This understanding, combined with simulations and empirical studies, likely led to the exploration and adoption of the Koch fractal structure for UWB-antenna applications.

The Koch fractal is very popular among researchers, but it also has some limitations, such as the Koch fractal structure's intricate geometry creating several challenges during the design and production stages. This is especially true when trying to achieve exact dimensions and iterations, especially when producing large quantities of the structure. Although fractal antennas have wideband capabilities, the Koch fractal has bandwidth limitations, making it difficult to attain wide frequency coverage without sacrificing other crucial antenna characteristics. Furthermore, matching the antenna's impedance to the system's impedance across its operating bandwidth is extremely difficult due to the complicated network of resonances and erratic patterns inherent in the Koch fractal structure. The more fractal iterations there are, the harder it is to shape the radiation pattern, which makes the directivity and radiation of the antenna less predictable. The proposed antenna has used the microstrip line for better performance. There are various ways to feed the antenna, one of which is CPW. CPW reduces the overall size of the antenna, but the reasons for choosing the microstrip feed line instead of CPW are various advantages of the microstrip line and the proposed design requirements, which can be stated as follows:

- i. Microstrip lines are easier and cheaper to fabricate using standard printed circuit board (PCB) manufacturing processes. This simplicity in fabrication makes them more cost-effective and practical for mass production.
- ii. Microstrip lines are easily integrated into standard PCB technologies, making them highly suitable for miniaturized and integrated designs.
- iii. In some cases, microstrip lines can exhibit lower dispersion compared to CPW. This characteristic is beneficial in maintaining signal integrity at higher frequencies. Microstrip lines allow for simpler integration of other passive and active components on the same substrate, facilitating a more compact and integrated antenna system.

Following a comprehensive literature review, some limitations were identified, which can be summarized as follows:

- i. Compliance with FCC regulations is a primary concern for ultra-wideband (UWB) technology. The UWB frequency range, typically spanning from 3.1 to 10.6 GHz in the United States, poses a considerable challenge in antenna design due to the need for effective operation across this wide frequency spectrum while managing issues related to size and performance.
- ii. The substantial bandwidth of UWB signals necessitates antennas that can efficiently cover this extensive frequency range while maintaining consistent impedance matching.
- iii. To maximize both the range and quality of UWB signals, antennas must consistently maintain a high level of radiation efficiency and a high gain.

The contributions of this paper are outlined as follows:

- i. Because of their self-similar structure, Koch fractal antennas may function across a range of frequencies, enabling multiband operation without requiring significant changes to the antenna's design. This antenna exhibits resonance at three distinct frequencies; namely, 3.3 GHz, 6 GHz and 8.6 GHz.
- ii. As per FCC, the bandwidth of the UWB antenna should be from 3.1 GHz to 10.6 GHz. The proposed antenna bandwidth spans between 2 GHz and 12.1 GHz, which covers the UWB range.
- iii. To optimize the range and quality of ultra-wideband (UWB) signals, antennas need to consistently uphold a superior level of radiation efficiency and a high gain. With the proposed antenna boasting a peak gain of 8.94 dBi and an efficiency of 96.8%, these attributes significantly bolster its position as a formidable candidate for the UWB frequency band.
- iv. The achieved three distinct resonance frequencies; namely, 3.3 GHz, 6 GHz and 8.6 GHz, are used in many different technological fields, including sensing, communication, radar systems and industrial applications and they take advantage of the special qualities that these frequency bands have to offer. Starting with the 3.3 GHz, within the microwave frequencies, these antennas power high-speed data transfers in 5G networks, enabling applications, like HD video streaming and swift mobile communication. The 6 GHz band, a cornerstone of Wi-Fi 6E, fuels seamless, high-speed connections in dense areas, such as stadiums and urban zones. Meanwhile, the 8.6 GHz frequency, part of the millimeter-wave spectrum, drives specialized high-speed, short-range communications, like WPAN and WBAN, as well as industrial, scientific and research applications. Together, these frequencies in ultraband antennas cater to a wide range of communication needs, balancing data rates, range and specific application requirements, thereby helping with precision in control systems, asset tracking and short-range sensing.

This paper starts by presenting an "introduction" that discusses fractal design concepts and challenges, followed by Sections 2 and 3 finalizing the Koch fractal design approach for fabrication. Patch-antenna design is carried out by a parametric study of the proposed UWB antenna. In Section 4, result analysis is conducted, followed in Section 5 by time domain analysis. Comparison with other reviewed references is shown in Section 6. This is followed by a "Conclusion", post-analysis of results and outcomes and suggestions for future work.

2. ANTENNA DESIGN APPROACH AND THEORY

The UWB antenna under consideration is constructed by implementing Koch fractal geometry along the edges of an octagonal shape. This antenna design represents a fusion of fractal geometries and conventional antenna technology. In fractal geometry, various segments are in different scales that come up again and again. Dimension and length [13] are derived using the following formulae:

$$D = \frac{\log(N)}{\log(r)} \quad (1)$$

$$l = h \left(\frac{N}{r} \right)^n \quad (2)$$

where, N= number of segments of geometry, h=length of curve, n = iteration number, r = number of segments divided by each iteration.

The algorithm of the iterative function system (IFS) is used to create the Koch fractal geometries. IFS is a very effective and useful method for generating any kind of fractal geometries. Translation, scaling and rotation are a few of the transformations used by IFS [10]. These changes are expressed as follows:

$$W \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \frac{1}{r} \cos\theta & \frac{-1}{s} \sin\theta \\ \frac{1}{r} \sin\theta & \frac{1}{s} \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} e \\ f \end{bmatrix} \quad (3)$$

where, scaling factor = r and s along the x and y axes, rotation angle= θ in the x-y plane and linear translations = e and f in the x axis and y axis, respectively.

Koch curve linear transformation matrices using IFS [14] are defined as follows:

$$W0 = \begin{bmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (4)$$

$$W1 = \begin{bmatrix} \frac{1}{3}\cos45^\circ & -\frac{1}{3}\sin45^\circ & \frac{1}{3} \\ \frac{1}{3}\sin45^\circ & \frac{1}{3}\cos45^\circ & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (5)$$

$$W2 = \begin{bmatrix} \frac{1}{3}\cos45^\circ & \frac{1}{3}\sin45^\circ & \frac{1}{2} \\ \frac{1}{3}\sin45^\circ & \frac{1}{3}\cos45^\circ & \frac{1}{2}\cos45^\circ \\ 0 & 0 & 1 \end{bmatrix} \quad (6)$$

$$W3 = \begin{bmatrix} \frac{1}{3} & 0 & \frac{2}{3} \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (7)$$

$$F = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 1 \end{bmatrix} \quad (8)$$

$$[xy_0 = W_0F, xy_1 = W_1F, xy_2 = W_2F, xy_3 = W_3F]$$

$$P = [xy_0 \quad xy_1 \quad xy_2 \quad xy_3] \quad (9)$$

where P is the all-coordinate point, which is generally used for fractal geometry in each segment. The first iteration of the Koch fractal divides the initial length into three equal segments and then, two equal segments are substituted for the middle segment, as depicted in Figure 1. The Koch fractal curve is generated by the repetition of each iteration on each and every segment.

Further, this Koch fractal is introduced in the substrate at the edge of the octagonal structure. Figure 2. depicts the center Koch fractal which is the subtraction at the center of the octagonal structure of the antenna. Step-wise structure evaluation is depicted in Figure 3.

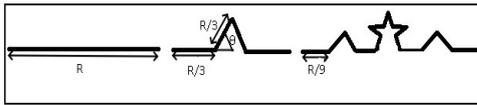


Figure 1. Generation of Koch fractal.



Figure 2. Centre Koch fractal.

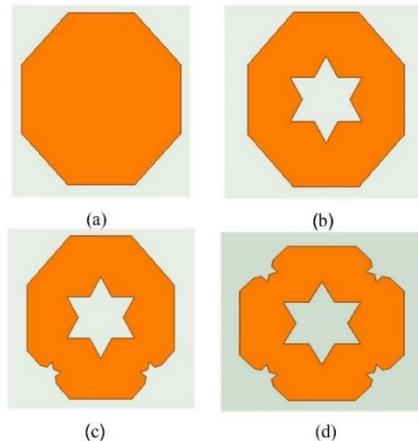


Figure 3. Koch fractal in octagonal geometry (a) Octagonal initiator, (b) First iteration, (c) Second iteration and (d) Third iteration.

3. INVESTIGATION OF PROPOSED ANTENNA ACROSS DIFFERENT PARAMETERS

In this antenna design, the octagonal structure serves as the initiator, while the generator employs the principles of the Koch fractal. The primary objectives include achieving compactness, miniaturization, wide bandwidth and a direction-independent design. The octagonal structure gives two advantages (a) in terms of area and circumferences, octagonal shape is close to a circular structure and (b) it's easy to apply fractals on the edge of the octagonal shape. This combination encourages the geometry's bigger circumference in miniature dimension in contrast to alternative geometries, such as hexagonal, elliptical, circular geometries and so forth. Each segment is in charge of providing a resonant frequency; so, when they come together, wideband behavior becomes apparent in design [15]. Also, the octagonal shape provides a uniform radiation pattern across a wide frequency range and provides compactness to reduce interference. In Figure 7(c), the comparison S parameter between the different diameters is shown. For designing the proposed antenna, θ is the beginning structure's embarking on length R, where convergence of fractal geometry is provided by θ . In this proposed antenna, $\theta=450$, $R=1.6$ mm for the edge of the octagonal iterative subtraction process and the center Koch $S2 = 5$ mm for the iterative subtraction process. It is evident that iteration of the Koch fractal in the octagonal generates the wide band and multiple resonant frequencies. Figure 4(a) depicts the progressive development stages of the proposed Koch fractal octagonal UWB antenna. Figure 4(b) shows the top view of the proposed antenna with measurement. Figure 4(c) depicts the bottom view of the proposed antenna with measurement. Figure 4(d) depicts the antenna with the microstrip feed line. Figure 4(e) shows the antenna with tapered feed.

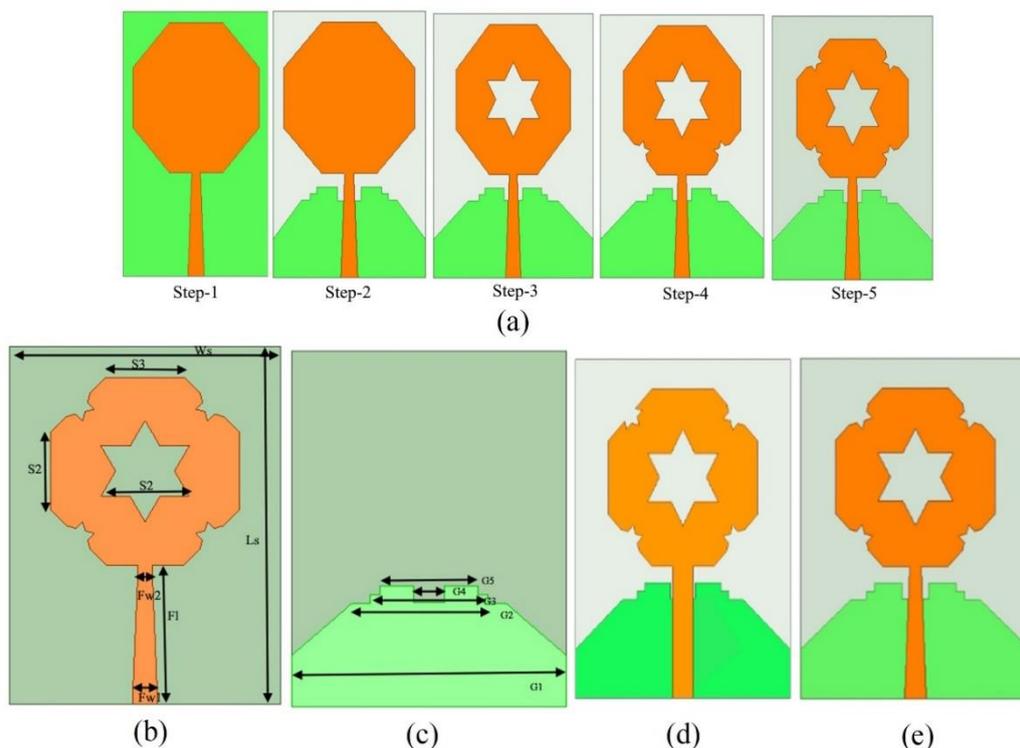


Figure 4(a) Progressive development stages of the proposed Koch fractal octagonal UWB antenna.

Figure 4(b) Top view of the proposed antenna with measurement. Figure 4(c) Bottom view of the proposed antenna with measurement. Figure 4(d) Antenna with the micro-strip feed line. Figure 4(e) Antenna with tapered feed.

The proposed antenna uses a defected ground and the main purpose of using DGS is its unique features. The antenna's resonance frequency may be impacted by DGS configurations. Multiband or broadband capability can be achieved by introducing certain patterns or structures that alter the operating frequency or create various resonance frequencies, which can assist in reducing losses, enhancing antenna performance and decreasing surface wave propagation on the ground plane's structures, improving the isolation between antennas and the overall performance of an antenna array by reducing mutual coupling

between elements when utilized in arrays. Closely-spaced antenna structures can improve the isolation between antennas and the overall performance of an array by reducing mutual coupling between elements [24]. The proposed antenna used a step-wise structure which helps in improving isolation and the overall performance. Figure 4(a) Step-1 is an antenna with no DGS and Step-2, Step-3 and Step-4 have come with DGS. The effect of DGS is shown in Figure 8(a).

The antenna design presented in this study is fine-tuned through optimization and its features are examined using HFSS software, version 21. Antenna design and analysis are carried out on FR4 substrate. Rectangular substrate works best for the basic UWB antenna, where the reason behind choosing the rectangular substrate is that it provides a wide band and good radiation property [16]. $\epsilon_r=4.4$ is the dielectric constant of the substrate and $\delta=0.02$ is the loss tangent of the substrate. The parameters of the optimized Koch fractal octagonal antenna are displayed in Table 1.

Table 1. Parameters of optimized Koch fractal octagonal UWB antenna.

Parameter	Dimension (mm)	Parameter	Dimension (mm)
L_S	34.3	Fw2	1.3
W_S	26.5	G1	26.5
S_2	5	G2	14.8
S_3	7.6	G3	10.4
h	1.6	G4	3.06
F_1	13.7	G5	9.4
FW_1	2.5	-	-

With the help of the basic formula of antenna-patch design [2], the overall length and breadth of the patch are calculated.

Effective Dielectric Constant:

$$\epsilon_{reff} = \frac{\epsilon_r+1}{2} + \frac{\epsilon_r-1}{2} \left[1 + 12 \frac{h}{w}\right]^{-2} \quad (10)$$

Width of patch:

$$W = \frac{1}{2f_r \sqrt{\mu_0 \epsilon_0}} \sqrt{\frac{2}{\epsilon_r+1}} = \frac{v_0}{2f_r} \sqrt{\frac{2}{\epsilon_r+1}} \quad (11)$$

Effective length of patch:

$$\frac{\Delta L}{h} = 0.412 \frac{(\epsilon_{reff}+0.3)\left(\frac{W}{h}+0.264\right)}{(\epsilon_{reff}-0.258)\left(\frac{W}{h}+0.8\right)} \quad (12)$$

$$L_{eff} = L + 2\Delta L \quad (13)$$

Resonance frequency of Antenna:

$$(f_r)_{010} = \frac{1}{2L\sqrt{\epsilon_r}\sqrt{\mu_0\epsilon_0}} = \frac{v_0}{2L\sqrt{\epsilon_r}} \quad (14)$$

The manufactured prototypes of the UWB antenna with a Koch fractal octagonal shape are produced and their attributes are verified within an anechoic chamber employing a Vector Network Analyzer (VNA). Figure 5 displays both the top and bottom views of the fabricated Koch fractal octagonal UWB antenna.



Figure 5. Top and bottom view of fabricated Koch fractal octagonal UWB antenna.

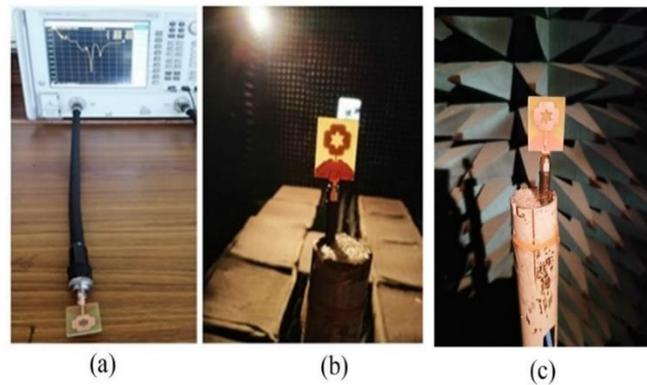


Figure 6. (a) Antenna under testing, (b) and (c) Antenna in anechoic chamber.

Following the fabrication process, the antenna design is subjected to laboratory testing, as illustrated in Figure 6. (a) Antenna under testing, (b) and (c) Antenna in anechoic chamber.

A parametric study of the antenna under consideration is conducted, which provides detailed information as well as the characteristics of the antenna. In ultrawide band antennas, geometrical and electrical parameters are affected by the desired characteristics. The main parameters of UWB antennas are dielectric constant of the used material (substrate), patch dimensions, feed gap, iteration geometric variation and dimensions of ground plane.

Figure 8 provides a sequential breakdown of the assessment of the proposed antenna. Step 1 illustrates the initial octagonal shape, Step 2 reveals the star Koch fractal subtraction in the center of the octagonal patch, Step 3 showcases the first two iterations on the lower side of the antenna and Step 4 displays the subsequent two iterations on the upper side of the recommended antenna. The ground plane for the proposed antenna is structured in a stepped shape, with its dimensions specified in Table 1.

The feed of the antenna used in the proposed design is in a tapered shape, which provides good impedance matching. From the literature survey, it is evident that for effective power transfer in the transmission line, which transmits the signal from the transmitter and the antenna itself, impedance matching is essential in antenna design. There are various ways that a tapered feed or tapered transmission line might help with impedance matching in antennas. The impedance of the transmission line and the antenna might gradually change in response to a tapering feed. The flow of power between the two is enhanced and reflections are decreased by this slow transition. By guaranteeing a smooth impedance transition, a tapered feed helps minimize standing waves by lowering reflections and increasing the amount of power that the antenna radiates. It facilitates improved impedance matching across a greater frequency range, which improves the antenna's performance. Figure 4(d) shows the antenna with a micro-strip feed line, while Figure 4(e) depicts the proposed antenna with tapered feed and the S parameter comparison graph is shown in Figure 7(b).

4. RESULTS AND ANALYSIS

All the results are analyzed in the lab and it is observed that all the results are up to the mark for the UWB antenna, as elaborated in the following sub-sections.

4.1 S Parameter

Figure 7(a) shows the stepwise S parameter of Koch fractal octagonal UWB antenna. Figure 7(b) compares S parameter between microstrip feed line and tapered feed line. Figure 7(c) Compares S parameter between different diameters of the radiator. Figure 7(d) compares S parameter between different heights of the substrate and shows the stepwise S parameter of the novel UWB Koch fractal antenna. From Figure 7(a), we can see in Step-1 the antenna having full ground not using DGS technology; the antenna does not get the desired bandwidth. In Step-2, the octagonal patch with DGS gets one resonant frequency at 3GHz and the bandwidth is narrow. After that, in Step-3, star Koch subtraction in the center of the octagonal patch improves the bandwidth of the antenna and resonates at 3.7 GHz. In Step-4, the first two iterations at the lower side of the proposed antenna provide the wide bandwidth and two resonant frequencies; 3.2GHz and 8.6 GHz. In Step-5, the second two iterations at the upper side of the proposed antenna increase bandwidth

covering from 2 GHz to 12.1 GHz. This antenna is efficient at resonating at the triple frequencies of 3.3 GHz, 6 GHz and 8.6 GHz. Figure 7(b) shows S-parameter graph between microstrip feed line and tapered feed line. From the graph, it is evident that with the microstrip feed line, the proposed antenna is not getting the UWB bandwidth, but with the tapered feed the proposed antenna gets the desired UWB bandwidth 2 GHz -12.1 GHz with three resonant frequencies. Figure 7(c) shows the comparison graph of the different diameters of the patch. From the graph, it is evident that the optimized diameter is helpful for the proposed antenna for gating the UWB bandwidth. Figure 7(e) shows the S-parameter comparison graph for different heights of the substrate, where it is observed that the 1.6 height is helpful for the proposed antenna.

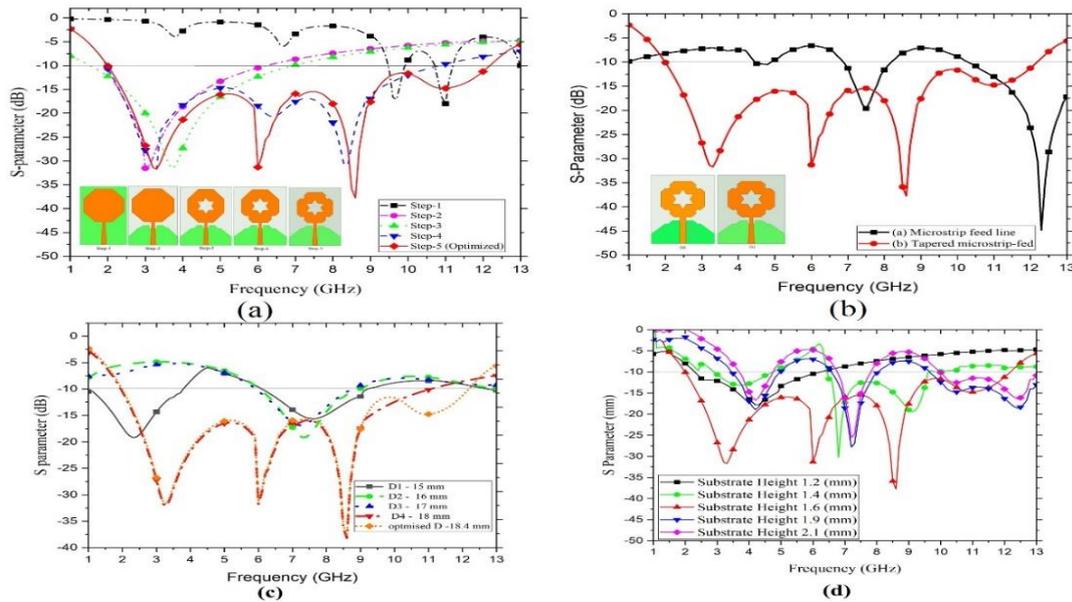


Figure 7. (a) The stepwise S parameter of Koch fractal octagonal UWB antenna, (b) S-parameter comparison between microstrip feed line and tapered feed line, (c) S-parameter comparison between different diameters of the radiator, (d) S-parameter comparison between different heights of the substrate.

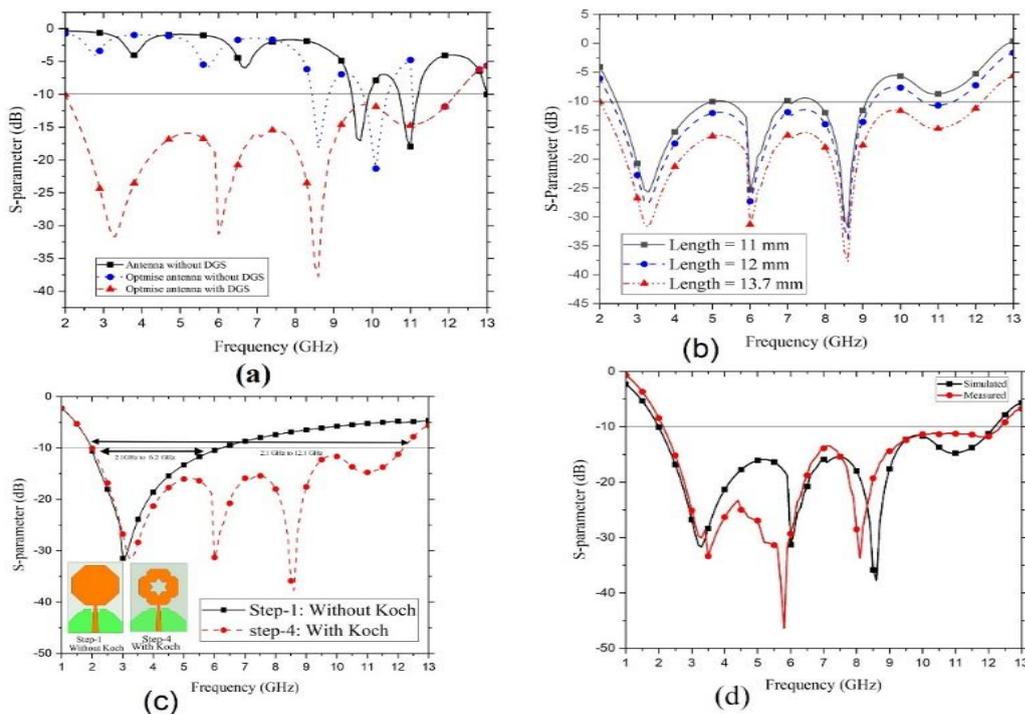


Figure 8. (a) S-parameter comparison with and without DGS, (b) S-parameter comparison of different feed lengths, (c) S-parameter comparison without Koch fractal and with Koch fractal octagonal UWB antenna, (d) Simulated and measured S parameters of the proposed antenna.

Figure 8 represents the parametrical study of the proposed antenna. Figure 8(a) shows the comparison graph of the proposed antenna, where in the base of the ground from the graph it is clear that the Step-1 antenna without DGS does not get the desired bandwidth. The optimized antenna without DGS also does not get the desired properties of UWB. Step-4 antenna optimized the proposed properties with DGS help for the desired bandwidth. Figure 8(b) shows the comparison graph of the different feed lengths of the proposed antenna and Figure 8(c) shows the comparison graph effect of the Koch fractal (with and without) in the proposed antenna, while Figure 8(d) depicts simulated and measured S parameters of the proposed antenna.

Here, we observe that the S parameter of the proposed antenna is below -10 dB and the triple resonance frequencies obtained are 3.3 GHz, 6 GHz and 8.6 GHz. It is observed that because of losses created by the SMA connector, the measurement surrounding and the fabrication tolerance, some discrepancies are shown in the results.

4.2 Radiation Pattern

Simulated and measured E-plane and H-plane radiation patterns at the resonant frequencies 2 GHz and 3.3 GHz of Koch fractal octagonal UWB antenna are shown in Figure 9. We observe that the radiation pattern at both frequencies is omnidirectional. Figure 10 shows the simulated and measured E-plane and H-plane radiation patterns at the resonant frequencies of 6 GHz and 8.6 GHz. Figure 11 shows the simulated and measured (a) E-plane and (b) H-plane radiation patterns at the resonant frequency of 12 GHz. Here, the three main resonant frequencies the beginning 2 GHz, the center 6 GHz and the end 12.1 GHz as well as the frequencies 3.3 GHz and 8.6 GHz are taken for the observation of the radiation pattern and Figure 11(c) shows the impedance matching on the Smith chart of the proposed antenna. Achieving perfect impedance matching across the entire UWB spectrum might be challenging due to the wide frequency range. From the graph, it is evident that the impedance matching of the proposed antenna is acceptable.

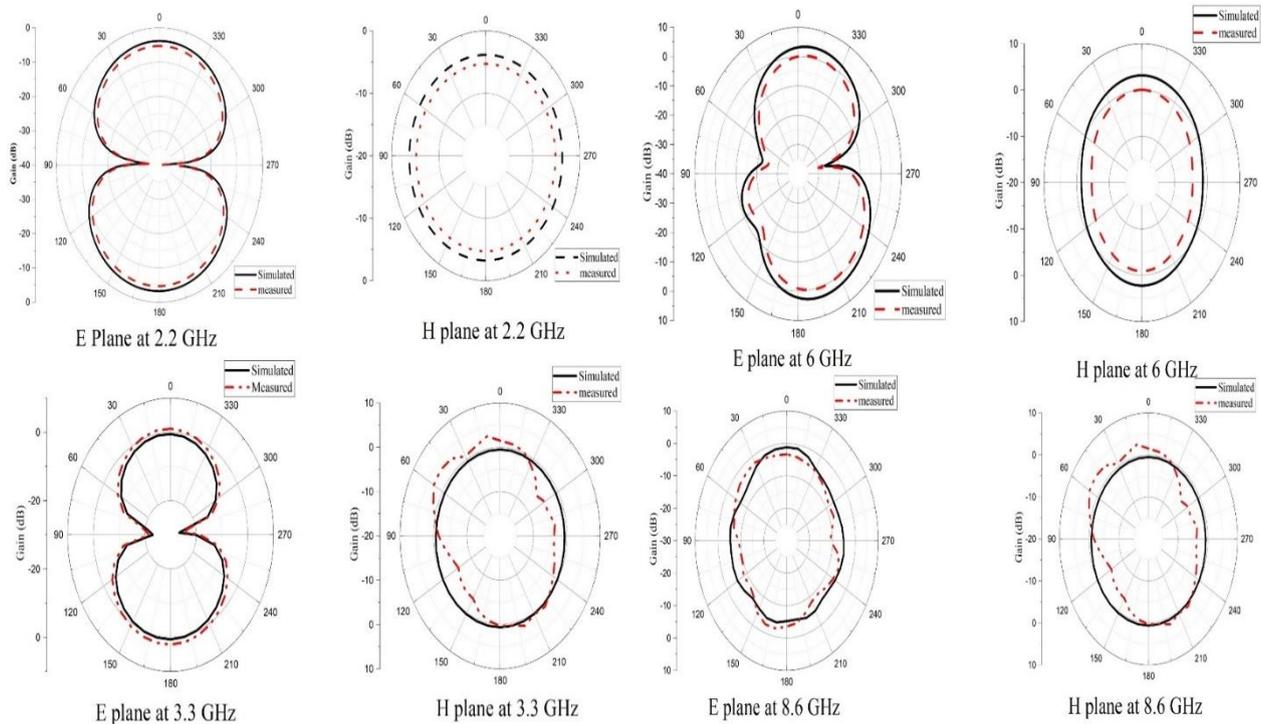


Figure 9. Simulated and measured radiation pattern in the E plane and H plane at 2.2 GHz and in the E plane and H plane at 3.3 GHz of the proposed antenna.

Figure 10. Simulated and measured radiation pattern in the E plane and H plane at 6 GHz and in the E plane and H plane at 8.6 GHz of the proposed antenna.

We observe that the antenna is efficient in resonating at frequencies of 2 GHz, 3.3 GHz, 6 GHz and 8.6 GHz. In the case of E-plane and H-plane radiation patterns, it is evident that in both planes, the radiation pattern is omnidirectional.

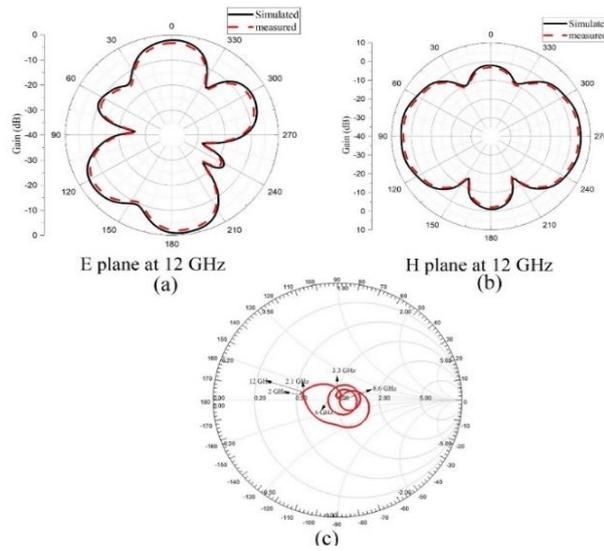


Figure 11. Simulated and measured radiation patterns (a) E-plane radiation patterns and (b) H-plane radiation patterns at 12 GHz of the proposed antenna and (c) Impedance matching on smith chart graph of the proposed antenna.

4.3 Radiation Efficiency

Radiation efficiency is also analyzed in this paper. Figure 12 shows the simulation followed by measured radiation efficiency vs. frequency graph. It is observed that during the resonant frequencies of 3.3 GHz, 6 GHz and 8.6 GHz, the radiation efficiency is between 92 and 96%. It is also evident that simulated radiation efficiency and measured radiation efficiency are flat around 2 GHz - 5.5 GHz.

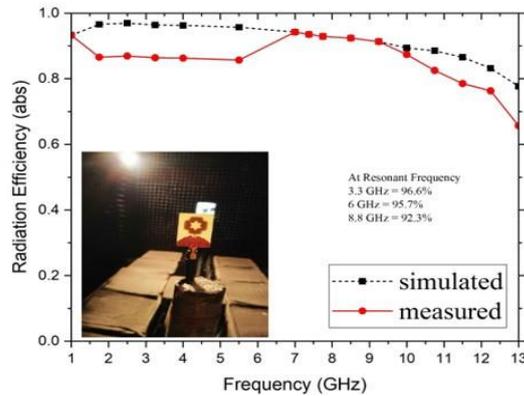


Figure 12. Radiation efficiency for the proposed Koch fractal octagonal UWB antenna.

4.4 Surface Current Distribution

Figure 13 and Figure 14 depict the Koch fractal antenna's surface-current distribution at resonance

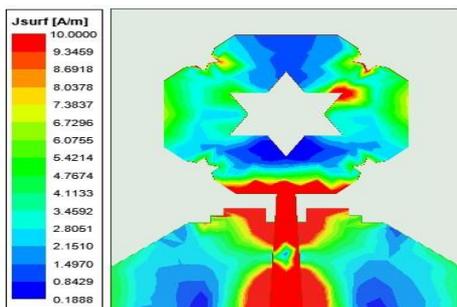


Figure 13. Surface-current distribution at resonant frequency of 3.3 GHz of the proposed Koch fractal octagonal UWB antenna.

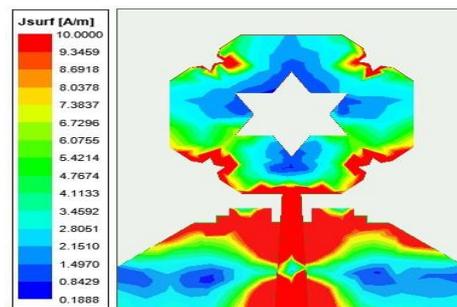


Figure 14. Surface-current distribution at resonant frequency of 8.6 GHz of the proposed Koch fractal octagonal UWB antenna.

frequencies of 3.3 GHz and 8.6 GHz. It is observed that at 3.3 GHz, the current is concentrated near the feed, but at 8.6 GHz, the surface current gathers surrounding the feed and the first half of the Koch antenna, which shows the accomplishment of the optimum impedance matching.

4.5 Gain

Figure 15 depicts simulated and measured gain of the Koch fractal octagonal antenna. Here, it was observed that the maximal gain of the designed antenna is 6.79 dB at 12.25 GHz. The overall gain is up to the mark for the UWB antenna.

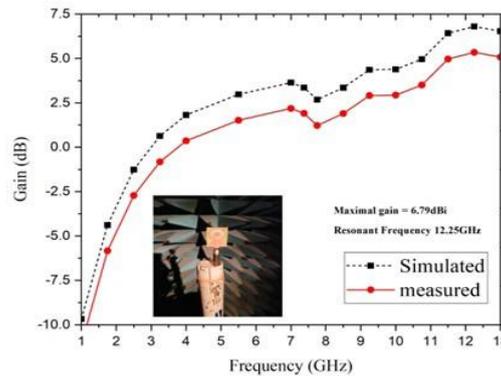


Figure 15. Peak gain vs. frequency graph of proposed Koch fractal octagonal UWB antenna.

5. TIME DOMAIN ANALYSIS

In the frequency domain, the antenna system demonstrates distinct properties, such as gain, polarization and directivity. Yet, examining the time domain provides the precise performance requirements for ultra-wideband antennas. This study sets important standards for impulse-oriented systems and confirms that they meet the requirements of the intended application. Time-domain characteristics, such as group delay, input and output correlation coefficients and fidelity-factor analysis, are explained in this section.

The correlation coefficient signifies the maximum correlation between two signals at a time delay τ , indicating the resemblance between the transmitted and received pulses. A maximum correlation coefficient of 0.9 denotes an almost complete similarity between the received and input signals of the proposed antenna [25]. The mathematical expression for the time correlation coefficient is calculated with the help of Equation 15.

$$F = \max \frac{\int_{-\infty}^{+\infty} s_{tx}(t)s_{tx}(t-\tau)dt}{\sqrt{\int_{-\infty}^{+\infty} |s_{tx}(t)|^2 dt \int_{-\infty}^{+\infty} |s_{rx}(t)|^2 dt}} \quad (15)$$

where, S_{tx} = transmitter, S_{rx} = receiver and τ = time delay.

Antennas with a greater fidelity factor provide transmitted signals with less distortion and better precision. This is closely related to having a larger bandwidth, which makes it possible to send and receive a wider range of frequencies efficiently. High-fidelity factor antennas are essential for maintaining clear and dependable signal transmission in applications like microwave imaging, where accuracy and precision are critical. This eventually improves the imaging-process quality and resolution [26]. The transmitted (T_x) and received (R_x) pulses, designated as T^n and R^n , respectively, are normalized using Equations 16 and 17. To compare the pulses' waveforms or shapes rather than their absolute amplitudes, this normalization step is necessary. To emphasize the shapes or patterns of the pulses rather than their magnitudes, the pulses are normalized, particularly when the R_x pulses' amplitude is smaller than the T_x pulses' magnitude. Equation 18 is used to find the fidelity factor. In the face-to-face configuration, the proposed antenna has a higher fidelity factor of 90.2; so, it is evident that the proposed antenna has less distortion and better precision.

$$T_s^n = \frac{T_s(t)}{\sqrt{\int_{-\infty}^{+\infty} |T_s(t)|^2 dt}} \quad (16)$$

$$R_s^n = \frac{R_s(t)}{\sqrt{\int_{-\infty}^{+\infty} |R_s(t)|^2 dt}} \quad (17)$$

$$FF = \max \int_{-\infty}^{+\infty} T_s^n(t) R_s^n(t)(t + \tau) dt \quad (18)$$

Another crucial performance metric for ultra wideband antennas is the group delay, which signifies the alteration in the phase of the transmitted signal concerning the frequency. In ultra-wideband (UWB) antennas, group delay signifies the time that it takes for diverse frequency parts within the transmitted signal to traverse the antenna or system. It's vital for UWB antennas spanning broad frequencies, revealing how these frequency elements experience delays within the antenna setup. Consistent group delay is crucial in UWB applications, as it upholds the timing relationships among frequencies, ensuring that these signal components reach their destinations without distortion or phase issues and maintaining signal accuracy and integrity.

In ultra-wideband (UWB) antennas, the group delay can be mathematically represented by deriving the phase response concerning frequency. If the phase response $\phi(\omega)$ of the antenna system is known as a function of angular frequency ω , the group delay τ can be computed by taking the negative derivative of the phase concerning angular frequency. Group delay is calculated with the help of Equation 19.

$$\tau = \frac{\delta\theta(\omega)}{\delta\omega} \quad (19)$$

Here, τ =group delay, $\phi(\omega)$ =phase response of the antenna system and ω = angular frequency.

To compute the correlation coefficient, we utilize a wideband Gaussian pulse as the input signal transmitted through the antenna, as demonstrated in Figure 16 The transmitter and receiver antennas are positioned 17 cm apart at various angles ϕ (0° , 45° , 90°) in the XZ plane. To find out the separating distance of the transmitter and receiver antennas, Equation 20 is used.

$$FarField \geq \frac{2D^2}{\lambda} \quad (20)$$

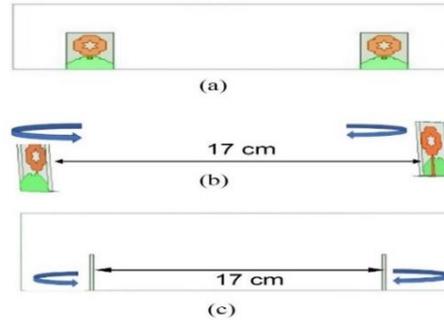


Figure 16. The transmitter antenna and receiver antenna are positioned 17 cm apart on the XZ plane (a-c), at various angles ϕ (0° , 45° , 90°).

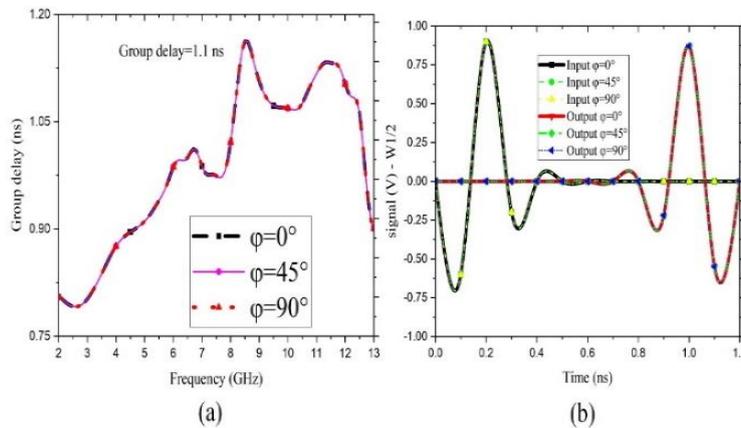


Figure 17. Time-domain characteristic graph (a) group-delay graph for $\phi(0^\circ, 45^\circ, 90^\circ)$ on the XZ plane at various angles ϕ ($0^\circ, 45^\circ, 90^\circ$) (b) graphs depicting the correlation coefficient.

Figure 17(a) shows the group-delay graph at various angles ϕ ($0, 45, 90$) degrees. In Figure 17(b), the graphs depicting the correlation coefficient are displayed for observation. The time-domain characteristics are acceptable for UWB. The group delay is 1.1 ns.

6. ANALYSIS IN RELATION WITH OTHER REVIEWED REFERENCES

The analysis of other reviewed references' antenna in terms of material used, dimension of antenna, operating band and gain are listed in Table 2. It has been noticed that the current work shows bandwidth enlargement. Moreover, the antenna size is smaller than in the works presented.

Table 2. Analysis of relations with other antennas.

Ref. No.	Material Used	Antenna Dimension mm ³	Operating Band (GHz)	Gain (dBi)
[9]	FR4	23.5x26.5x6	1.4-10.4	NR
[10]	RT/Duroid 5880	16.5x13.5x0.787	3.1-10.6	NR
[12]	Taconic	67x84x1	0.5-13.5	3.5-7
[13]	FR4	23.5x15x1.6	2.3-14	6.3
[16]	FR4	40x1.6x1.6	2-10	NR
[17]	FR4	30x30x1.6	2.3-13.2	5.5
[18]	FR4	60x55x1.59	NR	7.37
[19]	FR4	18.5x9.2x1.6	4.3-15.5	7.37
[20]	Taconic	30x40x0.8	3.3-11	5.74
[22]	FR4	40x24.5x1.6	NR	9.02
[23]	RT/Duroid 5880	24x30x0.787	3-12.7	6.1
[24]	Roger 5880	35 × 35 × 1.57	3.08-40.9	5.9
[25]	Plexi-glass	29× 27 × 1.6	2-32	10.6
[26]	RO5880 RF teflon	26×29 ×0.787	3.8-12.7	3.5
[27]	RO5880	52×26 ×0.787	2.3-11.5	8.4
[28]	FR4	14× 18 × 1.6	3.3-11.5	1.4
[29]	FR4	49.35x98.7x1.6	1-4.4	NR
[30]	RO4003C	71.6x71.6x3.118	3.27-3.82	8.1
[PW]	FR4	34.3 x 26.5 x 1.6	2-12.1	8.94

[*PW=proposed work,*NR = not reported].

Hence, it can be concluded that the suggested design of a single-element UWB antenna utilizing the Koch fractal octagonal structure is well-suited for UWB applications.

7. CONCLUSION

In this paper, a Koch fractal octagonal antenna is proposed for the ultra wideband (UWB) frequency-range applications. The uniqueness of this paper resides in its innovative use of the Koch fractal octagonal geometry to design an ultra wideband antenna that achieves miniaturization, compactness, wideband operation, high efficiency and multiple resonances. The combination of these features makes it a promising candidate for various wireless communication applications.

The paper introduces an innovative antenna design that utilizes a Koch fractal octagonal geometry. This choice of geometry is not commonly seen in traditional antenna designs, which makes it an innovative approach. The Koch-fractal use in the antenna design serves a dual purpose. It aids in miniaturization, meaning that the antenna can be made smaller while maintaining or enhancing its performance. Additionally, the fractal geometry contributes to the compactness of the antenna.

The proposed antenna's resonance at three specific frequencies (3.3 GHz, 6 GHz and 8.6 GHz) is another notable feature. This could potentially find applications in multi-frequency communication systems or frequency-selective applications. The antenna design presented provides a substantial bandwidth spanning from 2 GHz to 12.1 GHz. This kind of wide bandwidth is essential for UWB applications, This necessitates the capacity to send and receive signals across a wide spectrum of frequencies. The maximum gain of the antenna is 8.94 dBi at 12.25 GHz. The overall efficiency of the antenna is 96.8%. The time-domain characteristics are acceptable for UWB, where the group delay is

1.1 ns. The proposed antenna has a high fidelity factor of 90.2 and the correlation coefficient is 0.9, which makes the antenna a good candidate for UWB.

The paper introduces the significance of antenna design in modern wireless communication systems. It briefly mentions the challenges and the need for antennas with multiband operation, compact size, expanded bandwidth, high gain and high efficiency. These are the advantages of the proposed antenna:

Multiband Operation: The self-similar structure of the proposed Koch fractal antenna enables seamless operation across a wide range of frequencies. This unique design characteristic facilitates multiband functionality without necessitating significant alterations to the antenna's original design. The inherent scalability and self-similarity empower the antenna to operate efficiently across various frequency bands, marking a significant advancement in antenna versatility.

Compact Size: The fractal architecture of Koch-fractal antennas allows for remarkable miniaturization, rendering them ideal for integration into diminutive wireless devices, like RFID tags and smartphones. This inherent compactness without compromising performance positions these antennas as pivotal components for the next generation of portable communication technologies.

Expanded Bandwidth: Compared to conventional antennas, fractal antennas, including the proposed design, boast notably larger bandwidths. This expanded bandwidth offers unparalleled flexibility in signal transmission and reception, facilitating the accommodation of diverse data streams and improving overall communication reliability and performance.

High Gain and Efficiency: Maintaining consistently high efficiency throughout the ultra-wideband (UWB) spectrum, the proposed antenna stands out for its exceptional gain. The inherent design features optimize signal transmission and reception, resulting in amplified signal strength and reliability across the designated frequency range.

Summarizing the pivotal advantages presented by the proposed Koch fractal octagonal antenna, the proposed antenna has various applications. The bandwidth of 2 GHz to 12.1 GHz and the resonant frequencies of 3.3 GHz, 6 GHz and 8.6 GHz of the proposed antenna are important in the context of ultra-wideband (UWB). These various frequency allotments serve a variety of purposes according to regulatory requirements, bandwidth requirements, range constraints and interference considerations. Such antennas are used in many different technological fields, including sensing, communication, radar systems and industrial applications and they take advantage of the special qualities that these frequency bands have to offer. Starting with the 3.3 GHz, within the microwave frequencies, these antennas power high-speed data transfers in 5G networks, enabling applications, like HD video streaming and swift mobile communication. The 6 GHz band, a cornerstone of Wi-Fi 6E, fuels seamless, high-speed connections in dense areas, such as stadiums and urban zones. Meanwhile, the 8.6 GHz frequency, part of the millimeter-wave spectrum, drives specialized high-speed, short-range communications, like WPAN, WBAN, industrial, scientific and research applications. Together, these frequencies in ultra-band antennas cater to a wide range of communication needs, balancing data rates, range and specific application requirements, helping with precision in control systems, asset tracking and short-range sensing.

Future aspects of the proposed antenna are to improve the design for more applications in UWB, by designing the MIMO antenna, as well as the reconfigurable antenna.

REFERENCES

- [1] Revision of Part 15 of the Commissions' Rule Regarding Ultra-wideband Transmission System, First Report and Order, "Federal Communications Commission FCC 02-48," Washington, Apr. 2002.
- [2] C. A. Balanis, *Antenna Theory Analysis and Design*, 4th Edition, John Wiley & Sons, Inc., Hoboken, New Jersey, 2016.
- [3] D. H. Werner, R. L. Haupt and P. L. Werner, "Fractal Antenna Engineering: The Theory and Design of Fractal Antenna Arrays," *IEEE Antennas and Propagation Magazine*, vol. 41, no. 5, pp. 37-59, 1999.
- [4] J. Anguera, C. Puente, C. Borja and J. Soler, "Fractal-shaped Antennas: A Review," *Encyclopedia of RF and Microwave Engineering*, vol. 2, pp. 1620–1635, DOI: 10.1002/0471654507.eme128, 2005.
- [5] C. Borja and J. Romeu, "On the Behaviour of Koch Island Fractal Boundary Micro-strip Patch Antenna," *IEEE Transactions on Antennas and Propagation*, vol. 51, no. 6, pp. 1281–1291, 2003.

- [6] V. V. Reddy and N. V. S. N. Sarma, "Triband Circularly Polarized Koch Fractal Boundary Microstrip Antenna," *IEEE Antennas and Wireless Propagation Letters*, vol. 13, pp. 1057-1060, 2014.
- [7] M. R. Haji-Hashemi, M. M. M. Sadeghi and V. M. Moghtadai, "Space-filling Patch Antennas with CPW Feed," *Proc. of Progress in Electromagnetics Research Symposium*, pp. 26–29, Cambridge, USA, 2006.
- [8] M. G. Fekadu and S. N. Sinha, "UWB Fractal Slot Antenna Designs," *Proc. of the 2011 IEEE Int. Conf. on Microwaves, Communications, Antennas and Electronics Systems (COMCAS)*, pp. 1–4, DOI: 10.1109/COMCAS.2011.6105799, 2011.
- [9] S. Tripathi, A. Mohan and S. Yadav, "Ultra wideband (UWB) Antenna Using Minkowski Like Fractal Geometry," *Microwave and Optical Technology Letters*, vol. 56, no. 3, pp. 2273–2279, 2014.
- [10] S. Tripathi, A. Mohan and S. Yadav, "A Multi Notched Octagonal Shaped Fractal UWB Antenna," *Microwave and Optical Technology Letters*, vol. 56, no. 11, pp. 2469–2473, 2014.
- [11] C. Puente-Baliarda, J. Romeu, R. Pous and A. Cardama, "On the Behavior of the Sierpinski Multiband Fractal Antenna," *IEEE Transactions on Antennas and Propagation*, vol. 46, no. 4, pp. 517-524, DOI: 10.1109/8.664115, April 1998.
- [12] D. Li and J.-f. Mao, "A Koch-like Sided Fractal Bow-tie Dipole Antenna," *IEEE Transactions on Antennas and Propagation*, vol. 60, no. 5, pp. 2242-2251, DOI: 10.1109/TAP.2012.2189719, May 2012.
- [13] S. Tripathi, A. Mohan and S. Yadav, "A Compact UWB Koch Fractal Antenna for UWB Antenna Array Applications," *Wireless Personal Communications*, vol. 92, pp. 1423–1442, 2017.
- [14] S. Dhar, R. Ghatak, B. Gupta and D. R. Poddar, "A Wideband Minkowski Fractal Dielectric Resonator Antenna," *IEEE Transactions on Antennas and Propagation*, vol. 61, no. 6, pp. 2895–2903, 2013.
- [15] N. Tasouji, J. Nourinia, C. Ghobadi and F. Tofigh, "A Novel Printed UWB Slot Antenna with Reconfigurable Band-notch Characteristics," *IEEE Antennas and Wireless Propagation Letters*, vol. 12, pp. 922–925, 2013.
- [16] A. Gobinath, N. Sureshkumar, K. K. Hema, T. Sureka and B. Pavithra, "Design of Koch Fractal Antenna for Wireless Applications," *International Journal of Engineering Research & Technology (IJERT) NCIECC-2017*, vol. 5, no. 9, 2017.
- [17] S. Mythili, K. Akshaya, B. Charanya and K. Ayyappan, "Design and Analysis of Koch Fractal Antenna for WLAN Applications," *ICTACT Journal on Microelectronics*, vol. 06, no. 02, pp.923-927, 2020.
- [18] M. Gupta and V. Mathur, "Koch Boundary on the Square Patch Microstrip Antenna for Ultra Wideband Applications," *Alexandria Engineering Journal*, vol. 57, no. 3, pp. 2113-2122, 2018.
- [19] S. Singhal, T. Goel and A. Kumar Singh, "Inner Tapered Tree-shaped Fractal Antenna for UWB Applications," *Microwave and Optical Technology Letters*, vol. 57, no. 3, pp. 559-567, 2015.
- [20] H. Z. Liu, J. C. Coetzee and K. Mouthaan, "UWB Antenna Array for Wireless Transmission along Corridors," *Microwave and Optical Technology Letters*, vol. 50, no. 4, pp. 886–890, 2008.
- [21] S. Mahesh, A. Khairnar and M. Hasan, "A New Approach to Fractal Antenna Design for UWB Applications: An Analysis," *Mathematical Statistician and Engineering Applications*, vol. 71, no. 4, pp. 2326-9865, 2022.
- [22] I. H. Nejadi et al., "UWB Circular Fractal Antenna with High Gain for Telecommunication Applications," *Sensors*, vol. 23, no. 8, p. 4172, 2023.
- [23] M. A. Khan, U. Rafique, H. Ş. Savci, A. N. Nordin, S. H. Kiani and S. M. Abbas, "Ultra-wideband Pentagonal Fractal Antenna with Stable Radiation Characteristics for Microwave Imaging Applications," *Electronics*, vol. 11, no. 13, p. 2061, 2022.
- [24] S. Ullah, C. Ruan, M. S. Sadiq, T. U. Haq, A. K. Fahad and W. He, "Super Wideband, Defected Ground Structure (DGS) and Stepped Meander Line Antenna for WLAN/ISM/WiMAX/UWB and other Wireless Communication Applications," *Sensors*, vol. 20, p. 1735, DOI: 10.3390/s20061735, 2020.
- [25] M. Rohaninezhad et al., "Design and Fabrication of a Super-wideband Transparent Antenna Implanted on a Solar Cell Substrate," *Scientific Reports*, vol. 13, no. 1, p. 9977, 2023.
- [26] H. Şerif Savcı, "A Four Element Stringray-shaped MIMO Antenna System for UWB Applications," *Micromachines*, vol. 14, no. 10, p. 1944, 2023.
- [27] S. Hassan Kiani et al., "An Ultra-Wide Band MIMO Antenna System with Enhanced Isolation for Microwave Imaging Applications," *Micromachines*, vol. 14, no. 9, p. 1732, 2023.
- [28] S. Hassan Kiani et al., "A Novel Shape Compact Antenna for Ultra Wideband Applications," *Int. Journal of Antennas and Propagation*, vol. 2021, p. 7004799, DOI: 10.1155/2021/7004799, 2021.
- [29] A. Chowdhury and P. Ranjan, "A Novel Circularly Polarized, CPW-based MIMO Antenna for 5G-Wireless Communication in Sub-6 GHz Band," *Int. Journal of Electronics Letters*, DOI: 10.1080/21681724.2023.2266608, 2023.
- [30] M. A. Sufian, N. Hussain, H. Askari, S. G. Park, K. S. Shin and N. Kim, "Isolation Enhancement of a Metasurface-based MIMO Antenna Using Slots and Shorting Pins," *IEEE Access*, vol. 9, pp. 73533-73543, 2021.

ملخص البحث:

يقدم هذا البحث هوائياً مُدمجاً ثماني الأضلاع مع بنية أساس مشوّهة مُصمّماً للاستخدام في الاتّصالات اللاسلكية في نطاق التّردّدات فائقة العرض (UWB). ويؤدّي التصميم المقترح إلى الحصول على حجمٍ صغيرٍ لهوائي مُدمجٍ ملائمٍ للعمل في نطاق التّردّدات فائقة العرض. يُستخدم الهوائي المقترح النّحاس كموادّة موصلة على طبقة أساس من نوع (FR-4) لها ثابت عزّل كهربائي يساوي (4.4) و ظلّ زاوية فُقد مقداره (0.02) و سُمك يبلغ (1.6) ملم.

يحقق الهوائي المقترح كسباً أقصى مقداره (8.94) ديسيبل عند تردّد مقداره (12.25) جيجاهيرتز، ويعمل في نطاق تردديّ يمتدّ من (2) جيجاهيرتز إلى (12.1) جيجاهيرتز. وللهوائي المصمّم في هذا البحث ثلاثة تردّدات رنين (3.3، و 6، و 8.6 جيجاهيرتز)، الأمر الذي يمنحه فعاليةً عاليةً، حيث تبلغ فعاليته الكلية (96.8%)، كما يبلغ معامل الارتباط له (0.9)، ممّا يجعله مرشّحاً جيداً للاستخدام في نطاق التّردّدات فائقة العرض (UWB).

والجدير بالذّكر أنّ خصائص الهوائي المقترح تؤهّله للاستخدام في شبكاتٍ لاسلكيةٍ متنوّعةٍ مثل (WPAN) و (WBAN)، كما يمتدّ استخدامه إلى التطبيقات الصناعيّة؛ إذ يُساعد بدقّةٍ كبيرةٍ في أنظمة التّحكّم وتتّبّع الأصول والاستشعار قصير المدى وأنظمة الرّادار.

CDRSHNET: VARIANCE-GUIDED MULTISCALE AND SELF-ATTENTION FUSION WITH HYBRID LOSS FUNCTION TO RESTORE TRAFFIC-SIGN IMAGES CAPTURED IN ADVERSE CONDITIONS

Milind Vijay Parse¹ and Dhanya Pramod²

(Received: 10-Nov.-2023, Revised: 2-Jan.-2024 and 23-Jan.-2024, Accepted: 25-Jan.-2024)

ABSTRACT

In challenging weather conditions, various visual impediments such as raindrops, shadows, haze and distortions from dirty camera lenses and codec errors adversely affect the quality of traffic-sign images. Existing methods struggle to address these issues comprehensively, necessitating an innovative approach to restoration. This paper introduces the Codec Dirty Rainy Shadow Haze Network (CDRSHNet) architecture, integrating self-attention (SA) and variance-guided multiscale attention (VGMA) mechanisms. SA captures global dependencies, enabling focused processing of relevant image regions, while VGMA emphasizes informative channels and spatial locations for enhanced representation. A hybrid loss function, combining Gradient Magnitude Similarity Deviation (GMSD) and Charbonnier loss, boosts image quality. When trained on a diverse dataset, CDRSHNet attains a remarkable 99.3% restoration accuracy, yielding an average SSIM of 0.978 and an average PSNR of 39.58 on the Real Image Dataset (RID). On the Synthetic Image Dataset (SID), the average SSIM is 0.963 and the average PSNR is 39.46. The proposed model significantly improves image clarity and facilitates precise interpretation.

KEYWORDS

Image restoration, Challenging weather conditions, Variance-guided multiscale attention, Custom loss function.

1. INTRODUCTION

Computer-vision tasks are pivotal in various applications, relying heavily on accurately interpreting visual data. However, these tasks encounter significant challenges when images are captured under adverse conditions or influenced by codec errors, rain, haze, shadows and blurry images due to dirty camera lenses. These adverse conditions introduce visual distortions, which diminish image quality and hinder visual elements' precise detection and classification. Real-world applications like automated driving systems and traffic-sign recognition systems require image restoration to recognize traffic symbols successfully. Concerning these challenges, traffic-sign image restoration becomes vital to enhance image quality and enable precise classification. Image restoration targets the improvement of degraded or corrupted images affected by factors, like noise, blur or compression, ultimately aiming to create high-quality images closely resembling the originals [2]-[3].

The usage of deep-learning methods for image restoration has gained significant popularity, as they can effectively learn to model the complex relationships between degraded images and their corresponding clean versions. Autoencoders, Convolutional Neural Networks (CNNs), Deep Residual Networks (ResNets), Recursive Neural Networks (RNNs) and Generative Adversarial Networks (GANs) are some of the popular image-restoration techniques [2]-[3].

This study proposes a deep learning-based approach for restoring traffic-sign symbols under five challenging conditions: images captured with a dirty camera lens, images captured under rainy and hazy environments, shadow-influenced images and codec error images. These issues reduce image sharpness, resulting in compromised contrast, color accuracy and blurs. Codec errors impact visual artifacts, like pixelation and color distortion. The proposed CDRSHNet incorporates a fusion of attention blocks and group normalization to enhance the quality of the restored images. Attention blocks selectively highlight essential features in the image, improving the efficiency of the network to distinguish between different types of traffic signs. Group normalization decreases the internal covariate shift, improving network stability and performance. Furthermore, we introduce a custom loss function that combines the

1. M. V. Parse is with Symbiosis International (Deemed University) (SIU), Pune, India. Email: parsemv@gmail.com
 2. D. Pramod is with Symbiosis Centre for Information Technology (SCIT), Symbiosis International (Deemed University), Pune, India. Email: director@scit.edu

Charbonnier loss and Gradient Magnitude Similarity Deviation (GMSD) to enhance the quality of the recovered images. The GMSD evaluates the structural similarity between the original and restored images, whereas the Charbonnier loss penalizes significant errors in the restored image. The proposed approach is assessed using the images in the CURE-TSR dataset. The performance of the proposed network is assessed by peak signal-to-noise ratio (PSNR), mean absolute error (MAE) and structural similarity index measure (SSIM). These metrics are extensively used in the image-processing community to gauge the quality of the restored images and their analysis provides a quantitative assessment of the recommended approach's effectiveness. The experiments showcase that the proposed method outperforms current practices in restoring traffic-sign images under challenging conditions. The findings underscore the potential of deep learning techniques to enhance object restoration in complex real-world scenarios.

The paper is structured into several sections. Section two presents a literature review of existing works focusing on image-restoration techniques for challenging conditions, such as codec error, dirty lenses, rain, haze and images with shadows. Section three discusses our proposed architecture, including attention blocks, group normalization and custom loss function. Section four provides an in-depth analysis of the results obtained from the model, including the model loss and accuracy for both the training and validation datasets. Additionally, this section reports on the efficacy of the proposed approach in restoring traffic-sign symbols in challenging conditions. The performance of the proposed model is then assessed in Section 5 using a variety of measures, including peak signal-to-noise ratio (PSNR), mean absolute error (MAE) and structural similarity index measure (SSIM). These metrics are extensively used in the image-processing community to gauge the quality of the restored images and their analysis provides a quantitative assessment of the recommended approach's effectiveness. Finally, Section 6 provides the concluding remarks on the research work.

1.1 Our Contributions

The primary objective of this study is to implement a novel deep learning-based method designed explicitly for restoring traffic-sign symbols in the presence of five challenging conditions. Our proposed CDRSHNet model utilizes a fusion of attention blocks and group normalization to improve the quality of restored images. The attention blocks selectively emphasize essential features, enhancing the network's ability to distinguish between traffic signs. Moreover, group normalization assists in mitigating internal covariate shifts, improving stability and overall performance. We introduce a custom loss function that combines the Charbonnier loss and Gradient Magnitude Similarity Deviation (GMSD) to further enhance image recovery. The GMSD evaluates the structural similarity between images, while the Charbonnier loss penalizes significant errors. Together, these components contribute to the overall improvement in image quality. In our evaluations of the CURE-TSR dataset, we employ metrics, such as peak signal-to-noise ratio (PSNR), mean absolute error (MAE) and structural similarity index measure (SSIM). These well-established metrics provide a quantitative assessment of the effectiveness of our approach, clearly demonstrating its superiority over existing methods in restoring traffic-sign images under challenging conditions.

2. LITERATURE REVIEW

Image restoration aims to restore the original image from a damaged or noisy image. Over the years, various techniques have been developed to address this problem, each with advantages and limitations [4]. Some commonly used image-restoration techniques are filtering, statistical and model-based approaches, such as the blind-deconvolution method, inverse filter, Wiener filter and constrained least squares filter. These methods can be categorized as linear or nonlinear and are designed to mitigate the effects of noise and blur in the image. The techniques aim to recover the original image from a degraded version by applying mathematical operations that enhance its quality. The method chosen depends on the image-restoration problem and the type of degradation present in the image [4]-[5].

Diffusion-based methods use partial differential and variational restoration technology to propagate known information to the region to be repaired. This method works well for small-scale image damage, but cannot handle large missing areas or complex textures. Texture-based methods estimate information on corrupted regions using texture features in the original image and filling in the missing part with the best matching block. These methods are more suitable for severe damage in the image and can quickly

recover the texture details of the damaged image regions [6]. Researchers also use regularization-based methods to solve image-restoration problems. These methods aim to find a solution that satisfies constraints or regularization terms. Prior knowledge or assumptions about the underlying structure of the image are used to define the regularization terms [7]. Recently, various deep learning-based techniques have shown considerable success in image restoration. These methods use neural networks (CNNs) to learn the mapping between degraded images and their corresponding clean images. By training on large datasets consisting of degraded and clean images, these networks can learn to restore images accurately. A detailed discussion on deep-learning approaches to image restoration is given in [2].

Along with single neural network-based methods, generative adversarial networks (GANs) have also been used for image-restoration tasks. A discriminator and a generator are the two neural networks that comprise GANs. The generator network produces a restored image as an output from a degraded image as input. On the other hand, the discriminator network aims to differentiate between the cleaned and restored images. GANs can learn to generate highly realistic restored images by training these networks together. Overall, image restoration has advanced significantly over the years, with techniques ranging from linear filtering to deep learning-based methods [8].

G. Kwon et al. (2017) introduced a new and challenging dataset and comprehensively evaluated existing deep learning-based and traditional machine learning-based algorithms on this dataset. The CURE-TSR dataset includes images captured in real-world and simulated environments, including challenging weather, lighting and occlusions. The dataset includes various image degradations, such as blur, noise, shadows and codec errors, making traffic-sign recognition more difficult. Experimental results confirm that the performance of traffic-sign recognition algorithms varies significantly under different types of image degradations, with some algorithms being more robust than others [1]. S. Ahmed, U. Kamal et al. (2021) recommended a modular framework to detect and recognize traffic signs under difficult weather conditions. The offered solution implements a CNN for traffic-sign detection and recognition (TSDR) with prior image enhancement, comprising four modules: a challenge classifier, Enhance-Net (an image-enhancement module), a sign detection CNN and a classification CNN. Enhance-Net is trained to enhance traffic-sign regions specifically, enabling accurate detection instead of the entire image. The image enhancement component uses an encoder-decoder CNN architecture to augment input image quality by removing various image degradations, such as noise, blur and rain. The enhanced image is then fed into the region-based CNN detector for traffic-sign detection. The region-based CNN detector uses a two-stage approach, where, during the first stage, a set of candidate regions are generated, which are used during the second stage to classify a traffic sign or background. The image-enhancement block comprises five sub-blocks: rain, snow, haze, dirty lens and lens-blur removal blocks. Each block is designed to address a specific type of challenge. Each sub-block is applied to the input image only if the challenge classifier detects the corresponding type of challenge [9].

R. Huang et al. (2018) put forward an autoencoder-based architecture for restoring compressed images corrupted with codec errors. The autoencoder is first trained on a large set of clean images to learn a prior distribution of image patches. During the restoration process, the compressed image is first decompressed and then the autoencoder is used to predict the corrupted pixels in the decompressed image caused by the codec errors. To achieve this, the corrupted image is first divided into small overlapping patches and each patch is inpainted separately using the autoencoder. The autoencoder-based inpainting method uses the prior knowledge of the autoencoder to predict the missing pixels in each patch. Then, the patches are merged to produce the final restored image [10]. S. Jeon, H. Kim et al. (2019) proposed a solution for restoring compressed images distorted by the compression process. The authors advocated for an autoencoder-regularization approach to restore the images to their original quality. The proposed method involves training an autoencoder network on a large dataset of uncompressed images. The trained network is then used to regularize the restoration process for compressed images by imposing a constraint on the restored image to be similar to the output of the encoder part of the autoencoder [11].

K. Zhang et al. (2020) addressed the problem of concealing errors that occur during the compression of video streams. The authors introduced a two-stage method that uses deep-learning techniques to enhance the reconstructed video frames. The suggested method uses a CNN model to estimate the missing information in the corrupted video frames. This estimation is then used to generate a "confidence map" that indicates the reliability of the estimation at each pixel location. The next stage uses another CNN

to refine the estimated frames based on the confidence map [12]. M. Uricar et al. (2021) introduced a method for detecting image staining due to a camera lens in self-driving scenarios using a data-augmentation approach using a GAN. The recommended method uses a CycleGAN architecture to generate synthetic images that simulate different levels of camera-lens soiling with the help of two pairs of generator and discriminator networks to learn the mapping between clean and soiled image domains. The first generator inputs a clean image and outputs a corresponding soiled image. In contrast, the second generator inputs a soiled image and outputs a corresponding clean image. The two discriminator networks help ensure that the generated images are realistic and belong to their respective domains. The synthetic images generated by CycleGAN are then used to augment the original dataset of clean and soiled images. The augmented dataset trains a convolutional neural network (CNN) classifier that can accurately detect camera-lens soiling in autonomous driving scenarios. The authors use a dataset of 10,000 images from a camera attached to the front of a car. The dataset contains clean and soiled images, with varying levels of soiling caused by rain, snow and mud [13].

X. Li, B. Zhang et al. (2021) presented a solution for removing artifacts caused by contaminants (such as dust, dirt and moisture) on camera lenses in videos captured by moving cameras. The solution consists of three stages: detection, localization and removal. In the recognition stage, a deep CNN is trained to detect the presence of contaminants in each video frame. In the localization stage, a motion-analysis method is used to estimate the movement of the contaminants in each frame. In the removal stage, using the estimated motion information, a temporal filtering method is applied to remove the artifacts caused by the contaminants. To evaluate the suggested approach, the authors created a dataset of 30 video sequences with varying levels of contaminants captured by a moving camera. The dataset includes videos with rain, snow, dirt and clean reference videos. The authors additionally offer annotated ground truth and the extent of the contaminants in each frame. The method achieves PSNR of 35.37 and SSIM of 0.980 in stage one [14]. J. Mohd, S. M. Reyes et al. (2021) described a novel approach to detect dust particles in camera lenses mounted on moving robots. The suggested method also includes a technique for correcting the recorded or live image data by selectively removing the dust areas using an adaptive tiling-based approach. Dust particles are a significant issue for camera lenses in different disciplines, such as traffic-sign identification and geospatial data capture. The method aims to improve dust detection and correct image data by comparing consecutive frames captured by moving robots and removing the dust particles using the proposed technique while preserving the original data. Simulation results achieved 90-92% accuracy in removing the dust particles without affecting the actual data, which is a significant improvement [15].

H. Wang et al. (2021) suggested a deep CNN-based solution, "SRNet," to remove rain from a single image, integrating a structural residual learning framework with a residual block and a multiscale structure-extraction network. The residual block is designed to distinguish between the input and output images, while the multiscale structure-extraction network is employed to extract structure information from input images. This method achieved satisfactory performance on rain removal in a single image with PSNR=35.31 and SSIM=0.9448 [16]. Rainy images can adversely impact multimedia and computer-vision applications. CNN-based solutions have been employed to address this issue and eliminate rain from a single image. S. Li, W. Ren et al. (2019) presented a novel multi-task learning architecture that enhances performance by reducing excessive mapping between ground truth and output images. This architecture features a decomposition network that separates the rainy image into a clear background and multiple layers for the main component. During training, the composition structure is reconstructed to enhance the image quality by integrating clean input images and rain-related information. Experimental results demonstrated that this approach produces high-quality image restoration for synthetic and real images and surpasses contemporary techniques. Furthermore, the technique can be applied to other tasks, such as dust abstraction. Their method achieves PSNR of 33.7508 and SSIM of 0.9412 on the Rain50 dataset [17].

By analyzing urban video scenes, vision-based traffic analytic systems can significantly benefit Intelligent Transportation Systems (ITSs). However, vehicle detection and tracking can be challenging due to moving cast shadows, resulting in inaccuracies. M. U. Arif et al. (2022) conducted a comprehensive analysis of traditional and cutting-edge shadow identification and removal algorithms for traffic scenes based on 70 research papers published over the past 30 years. The study emphasizes the need for a hybrid approach combining traditional and well-known shadow-detection and removal

techniques before applying CNN-based vehicle-detection methods. The study also recommends using Highway I, II and III datasets for comparative evaluations; despite many CNN-based techniques for vehicle detection, moving cast shadow is still a challenge, necessitating pre-processing steps for accurate vehicle detection in traffic scenes [18]. Eliminating shadows from images can improve their visual appeal and has numerous applications in computer vision. Currently, based on deep-learning techniques, CNN is deemed the most efficient way to remove shadows. These methods can be trained using paired data of both the shadowed and clean images. Training CNN on unpaired data is typically favored in practice owing to the simplicity of data collection. In 2021, Z. Liu et al. proposed a new approach to shadow removal, referred to as the Lightness-Guided Shadow Removal Network (LG-ShadowNet), which employs unpaired data for training. The method comprises two CNN modules, with the first compensating for lightness and the second removing the shadow based on the information obtained from the previous module. It also introduces a loss function that utilizes the color before the existing data. Comprehensive experiments were conducted on popular datasets, such as the Image Shadow Triplets Dataset (ISTD), adjusted ISTD and USR. The proposed method performs well compared to available methods trained on unpaired data, with PSNR=25.92 and SSIM=0.909 [19].

The limited availability of paired shadow and ground truth images hinders the development of robust and large-scale shadow extraction algorithms. This limitation restricts the variety and size of shadow-removal datasets, making it difficult to train such algorithms. H. V. Le et al. (2020) presented a new shadow-removal technique that uses shadow and non-shadow regions from images for training to address this challenge. The approach uses an adversarial framework that incorporates a physical shadow-formation model. The method is handy for video shadow removal and achieves good results compared to the existing works [20]. H. Fan et al. (2019) examined the difficulties that current shadow-removal methods face in image segmentation and target recognition. They recommended a deep CNN composed of an encoder-decoder and refinement network to address these issues. The network predicts the alpha shadow scale factor and generates sharper edge information. A new image database (RSDB) is built and tested against various databases to evaluate the algorithm. Compared to other methods, the suggested algorithm significantly improves PSNR and SSIM metrics, producing sharper and shadow-free images that retain the image's color and texture close to the original image [21]. X. Hu et al. (2019) introduced Mask-ShadowGAN, a novel technique for removing shadows using unpaired data. The approach uses a deep-learning framework to generate a shadow mask based on an input shadow image. The generated mask is then employed to guide the process of shadow generation, incorporating cycle-consistency constraints. The framework is designed to simultaneously learn the generation of shadow masks to optimize overall performance. The effectiveness of this approach was evaluated on an unpaired dataset for removing shadows; it exhibited promising results across various experiments [22].

A unique image-fusion technique was introduced in 2021 by L. Ren et al. It improves the guided filter for better decomposition and artifact reduction. Before fusion, the contrast of viewable pictures is improved to address low light and noise. The authors divide the visible and infrared images vertically into sub-images, separate them into base and detail layers and use two fusion techniques. They also suggest a gradient-brightness criterion for adaptive output. Compared to earlier fusion techniques, experimental results show more significant performance in maintaining visible image details and improving infrared object clarity [23].

Q. Yang et al. (2022) tackled the problem of small-object detection. Feature Pyramid Networks (FPNs) represent a revolutionary technique that the authors suggested for enhancing small item detection. Small-Object Feature Enhancement (SOFE) and Variance-guided Region of Interest Fusion (VRoIF) are the two modules that makeup SV-FPN. To extract small-object features, SOFE improves finer-resolution level feature maps. In addition, VRoIF uses the variation in RoI features to determine the degree to which various RoI characteristics from various layers are all present. Ablation tests demonstrate the efficiency of SV-FPN on three open datasets, which achieved mean Average Precision (mAP) values of 41.5%, 53.9% and 38.3% on the KITTI, PASCAL VOC 07+12 and MS COCO 2017 datasets, respectively [24]. X. Yang (2020) conducted a comprehensive review of the attention mechanisms in computer vision. It discusses several attention mechanisms: self-attention, channel attention and spatial attention. The author emphasizes using attention mechanisms in image synthesis, object detection and picture classification. The author also investigates various attention-incorporating models, such as Transformer-based models and Convolutional Neural Networks (CNNs) and evaluates their benefits and drawbacks. In computer vision, the importance of attention mechanisms in improving

visual perception and task performance is substantial [25].

DehazeFormer distinguishes itself by showcasing its superior performance across multiple datasets, highlighting its efficacy in image dehazing. The authors introduce several enhancements to the Swin Transformer architecture, incorporating modifications, such as replacing LayerNorm and Gaussian Error Linear Unit (GELU) with RescaleNorm and ReLU, respectively. Furthermore, they propose a shifted window-partitioning scheme and a spatial information-aggregation scheme, contributing to the model's resilience and efficiency in dealing with dehazing tasks. Importantly, these improvements transcend the scope of DehazeFormer, as they offer minor, yet impactful, enhancements that can be applied to other networks. The model's features are exceptionally noteworthy regarding its performance on a substantial remote-sensing image-dehazing dataset [26].

A task-related contrastive network for single-image dehazing is introduced by W. Yi et al. (2023). It focuses on a compact autoencoder-like architecture with FEM and AFM, utilizing contrastive learning for improved performance. The proposed strategy involves effective data augmentation and a task-friendly embedding network. TC-Net outperforms existing methods, but limitations include reliance on common data-augmentation approaches and increased GPU-memory usage due to dynamic parameter updates in the training process [27].

To summarize, image-restoration techniques encompass various methodologies targeting specific image flaws. Diffusion methods excel at rectifying minor damages, while texture-based approaches effectively recover severe image flaws. Regularization methods rely on prior knowledge to meet image-structure constraints for restoration. Deep-learning techniques employing Convolutional Neural Networks (CNNs) efficiently restore images by learning from extensive datasets. GANs, with their generator and discriminator, produce realistic image restorations. Autoencoder-based methods predict and restore corrupted pixels in decompressed images, while autoencoder regularization aids in restoring compressed images. Deep-learning techniques efficiently enhance reconstructed video frames post-compression errors. GAN-based augmentation effectively detects lens staining in autonomous driving scenarios. Techniques for detecting and removing camera-lens contaminants in moving videos follow systematic stages. Rain removal relies on CNN-based approaches using structural learning and residual networks. Multi-task CNN architectures efficiently handle shadow removal with minimal mapping. Hybrid methods combine pre-processing with CNN-based detection for shadow detection. Innovations, like Lightness-Guided ShadowNet, use unpaired data for effective shadow removal. Techniques involving physical shadow models and CNN-based refinement effectively address shadow-removal issues. Mask-ShadowGAN, utilizing unpaired data, showcases a comprehensive approach to shadow removal. Image-fusion techniques employ guided filters and adaptive output for improved clarity in restorations. Collectively, these techniques contribute to the broad landscape of image restoration, each offering unique solutions for diverse image imperfections.

3. PROPOSED CDRSHNET ARCHITECTURE

The CDRSHNet architecture accepts colored noisy images with a resolution of 128x128x3 (height, width, RGB channels) and outputs a restored noiseless image. The network design has five levels with a cascade succession of convolutional operations. On each level, there are residual blocks. Two convolutional layers with Rectified Linear Unit (ReLU) activation and a group-normalization layer comprise each residual block. Small batch-size issue is addressed using group normalization. The feature map is normalized using group normalization, which divides the channels into groups and generates unique normalization statistics for each group based on the channel mean and variance. This strategy ensures independence from batch size when batch sizes are small. Padding is used throughout the model during convolutional operations to preserve the edge information of the image. The feature maps' spatial dimensions are preserved through padding, allowing the convolutional layers to capture important spatial data effectively. A skip connection is formed by element-wise summing the output of each residual block with its corresponding input. By propagating significant features from earlier blocks to later ones, this skip link aids in preserving those features. Figure 1 shows the proposed architecture of CDRSHNet.

The output of the final block in a level is passed as a skip connection to the self-attention module after the execution of all three residual blocks. The self-attention module captures global dependencies, so

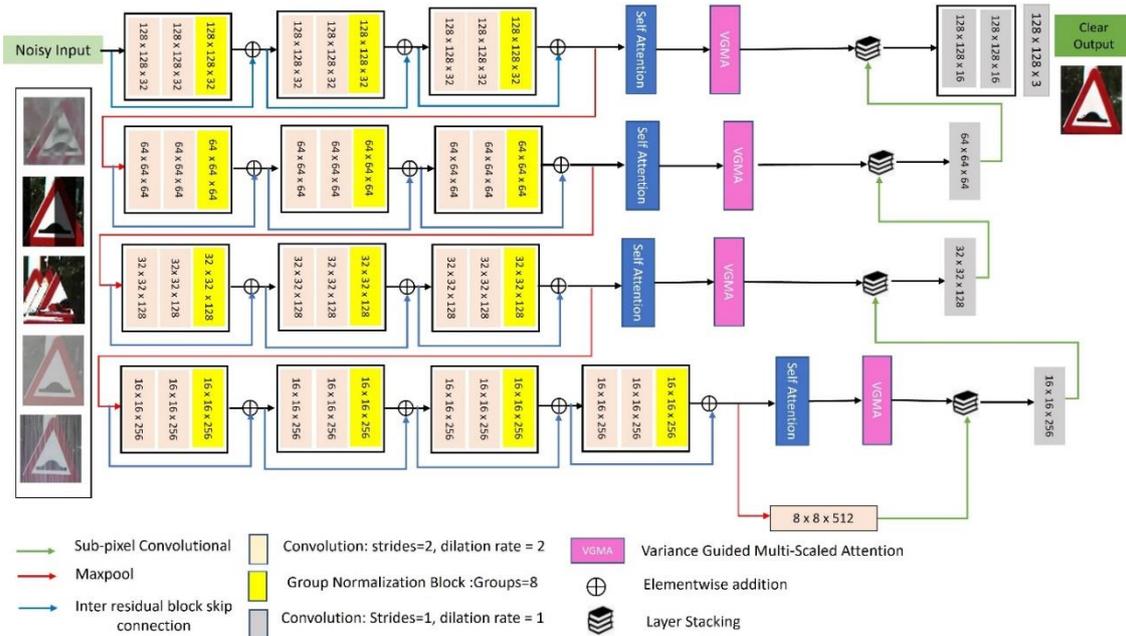


Figure 1. Proposed CDSRHNet model architecture.

that the network focuses on relevant image regions. The Variance Guided Multi-Scaled Attention (VGMA) module continues to process the output of the self-attention module and improves the image representation by emphasizing informative channels and spatial locations. Also, skip connections are used in deep neural networks to solve the vanishing-gradient problem, because the gradient signal has a shorter path to follow during backpropagation. The spatial dimensions of the feature maps produced at the end of the third residual block in each level are down-sampled using pooling layers. This downsampling reduces the parameter count in the network and increases its insensitivity to small spatial translations. Additionally, pooling layers increase the neurons' receptive field, which improves the network's ability to record spatially invariant data. Equation (1) determines the output down-sampled feature map of a tensor.

$$D = \frac{(I+2P-K)}{S} + 1 \quad (1)$$

where, I is the input feature map, P is the padding, K indicates convolutional kernel and S is the stride of convolutional operation.

The dilation process introduces gaps or spaces between the kernel elements, allowing neurons to perceive a larger area of the input-feature map. The dilation rate determines the size of the gap in the kernel. The proposed method has a dilation rate of (2,2), which increases the receptive field without increasing the parameters. The mathematical formula for a dilated convolutional operation with dilation rate d can be expressed as follows with the help of Equation (2).

$$y = \sum_{k=0}^{k-1} [i + (k * d)] * w[k] \quad (2)$$

where, i is input feature map, y is output feature map, w is the kernel of size k and d is the dilation rate.

The spatial dimensions of output-feature map y are identical to those of input-feature map i, but a factor of d expands each neuron's receptive field. The final bottleneck layer is 8x8x512 in dimension and represents a compressed-input representation. The most important and pertinent data from the preceding layers is collected in this layer. The expansion path uses subpixel convolution to improve efficiency, capture global dependencies and retain spatial information. The outputs from the VGMA module and the preceding level's subpixel convolution layer are combined at each level of the expansion path. Concatenation enables the model to recover spatial data that was lost during down-sampling.

3.1 Self-attention

The self-attention module, illustrated in Figure 2, captures long-range dependencies and selectively attends to relevant information within the input sequence or feature map. It requires three parallel 1x1 convolutions, to produce the query(Q), key(K) and value(V) vectors. While the key vectors represent all the other positions or elements, the query vector represents the current position or element being

attended to. The attention mechanism establishes position weights through the element-wise dot product of query and key vectors, followed by obtaining attention weights using a Softmax function.

The value vectors (V) containing features related to each position are multiplied by the respective attention weights. The self-attended feature representation is obtained by combining the resulting weighted value vectors using a summation operation or weighted average. The self-attended characteristics are then refined and transformed using a 1×1 convolution. Self-attention effectively captures the intra-correlation of an input matrix $X = [x_1, x_2, \dots, x_n] \in \mathbb{R}^{(d \times n)}$, where d is the dimensionality of the input vectors and n is the number of vectors in the sequence. In this self-attention, Q , K , V and X are kept equal and the self-attention is computed as in Equation (3).

$$SA = \text{Softmax} \left(\frac{X^T X}{\sqrt{d}} \right) * X \quad (3)$$

here, SA is the self-attended feature representation calculated based on the attention mechanism which captures the intra-correlation or dependencies between each pair of input vectors, reflecting the significance of each vector in understanding or representing other vectors in the sequence.

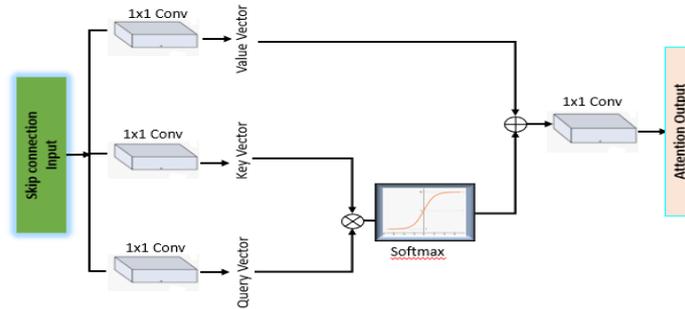


Figure 2. Self-attention architecture.

3.2 Variance-guided Multiscale Attention

The proposed approach combines channel attention with variance and spatial-attention modules to improve image-reconstruction quality while reducing network depth and computational complexity. By emphasizing channels with high variance, which indicates the presence of important or distinctive features, the channel-attention mechanism based on variance enables us to capture channel variability. The network can adjust its weighting based on variance by prioritizing informative channels and suppressing noisy or less informative channels. Each spatial position in the feature maps is taught to receive weight from the spatial-attention module, reflecting the significance of that region during feature fusion. This allows the network to adaptively adjust the contribution of each spatial location based on its significance, enhancing the representation of local and global details.

We perform an element-wise summation between the weighted feature maps obtained from the channel-attention mechanism and the spatial-attention weights to fuse the outputs of the channel attention and spatial-attention modules. By combining the informative channel-level weights with the spatially adaptable weights, this fusion process maximizes the benefits of both methods. This novel approach presents practical solutions to address the specific challenges outlined in this paper, offering advanced capabilities for various image-enhancement tasks. Figure 3 shows the architecture in more detail.

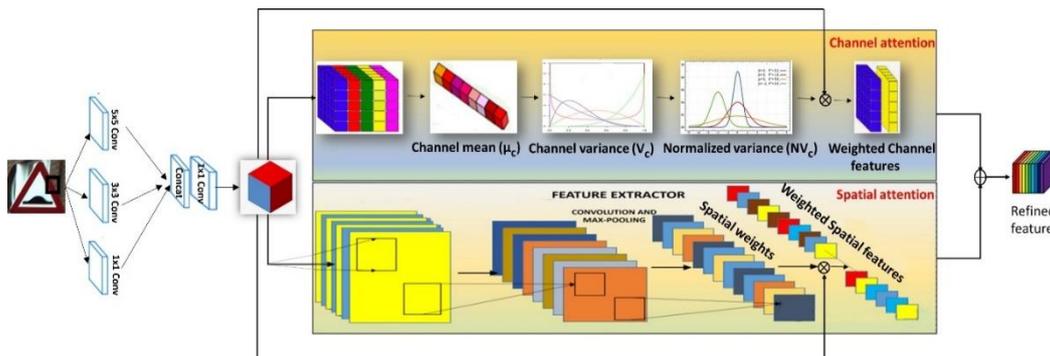


Figure 3. Variance-guided multiscale attention architecture.

3.2.1 Channel Attention

The channel-attention mechanism offers several benefits when variance is used to adjust the weights of all channels. First, emphasizing channels with high variance, which denotes the presence of significant or distinctive features, enables the capture of channel variability. The network can prioritize channels that contribute more to the overall task by using this adaptive weighting based on variance, which enhances performance. Incorporating variance-based weights further improves channel-noise robustness by minimizing the impact of noisy or uninformative low-variance channels. In addition to increasing the network's computing performance, this reduces noise and redundant data. The following are the steps involved in extracting significant channel-level features.

3.2.1.1 Variance Computation: The variance (V_c) for each channel (c) is calculated for a feature map (F) with dimensions $H \times W \times C$, where H represents height, W signifies width and C denotes the number of channels. The formula for V_c is expressed as follows:

$$V_c = \frac{1}{(H*W)} * \sum_{i=1}^H \sum_{j=1}^W (F(i, j, c) - \mu_c)^2 \quad (4)$$

Here, H and W denote the height and width of the feature map, respectively; i and j are indices representing spatial positions within the feature map; $F(i, j, c)$ is the value of the feature map at position (i, j) in channel c and μ_c represents the mean value of channel c along a particular dimension in the input-feature map.

3.2.1.2 Variance Normalization: Channel-specific variances are normalized to obtain weights that sum up to 1. The calculation of Normalized Variance (NV_c) is articulated as follows:

$$NV_c = \frac{V_c}{\sum V_c} \quad (5)$$

3.2.1.3 Weight Adjustment: To adjust the importance of each channel, normalized variances are applied as weights. The weight assigned to a channel increases with its variance. Equation (6) calculates the adjusted weighted feature map (W_{oc}) for spatial position (i, j) and channel c :

$$W_{oc}(i, j, c) = NV_c * F(i, j, c) \quad (6)$$

Each member of the matrix F is represented by $F(i, j, c)$, denoting the value at the position (i, j) in channel c of the feature map.

3.2.2 Spatial Attention

Integrating spatial attention with variance-based channel attention offers a comprehensive approach to identify spatial dependencies and channel-wise variability, improving feature representation. Spatial attention improves feature representation by considering the spatial dependencies within the feature maps, whereas variance-based channel attention concentrates on capturing inter-channel relationships and significant features. Spatial attention captures contextual information and fine-grained features by modeling the relationships between spatial locations and allocating weights accordingly. This leads to improved feature discrimination and improved model performance.

Consider $f(\cdot)$ being the function to compute the spatial-attention weights denoted by $SW(i,j)$, for an original feature map represented by $O(i,j)$ at each spatial location (i, j) . This involves a sequence of mathematical operations applied to the original feature map. After these operations, a nonlinear activation function is applied in conjunction with a weighted convolution. The spatial weights $SW(i,j)$ at spatial location (i,j) are given by Equation (7).

$$SW(i,j) = f(O(i,j)) \quad (7)$$

3.2.3 Adjusting Contribution by Feature Fusion

The weighted feature maps from the channel-attention mechanism created from Equation (6) are multiplied element-wise with the spatial-attention weights acquired from Equation (7) in the previous step to adjust the contribution of each spatial location during feature fusion.

$$W_{ocAdjusted}(i, j) = W_{oc}(i, j) * SW(i,j) \quad (8)$$

$W_{ocAdjusted}(i, j)$ in Equation (8) represents the adaptively adjusted channel weights at each spatial location (i,j) .

3.3 Hybrid Loss Function

This study proposes a novel custom loss function that combines the GMSD (Gradient Magnitude Similarity Deviation) and Charbonnier loss functions. Compared to the conventional single loss function, which can only handle one type of issue, the proposed loss function can simultaneously address noise, blurriness and sharpness problems. X. Zhu et al. [28] employed an integrated loss function made up of MSE, VGG19-based perceptual loss and Novel Quality Loss inspired by IQA (Image Quality Assessment) metric and GMSD to create an image that is in line with human vision, utilizing the GAN network.

The Charbonnier loss and the VGG16 perceptual loss were both employed by B. Wu et al. [29]. Their results outperformed previous-study methods in producing images with intricate features and sharp edges. In order to reduce noise in CT-scan images, B. Gajera et al. [30] proposed an enhanced GAN that combined Charbonnier loss and VGG19 perceptual loss. The authors found that this network significantly improves denoising performance by bringing soft-tissue noise levels closer to those of a Normal Dose CT (NDCT) scan. In Equation (9), L_M shows the final loss function of the proposed model. The weights for each loss term are α and β , respectively.

$$L_M = \alpha L_{\text{GMSD}} + \beta L_{\text{Charbonnier}} \quad (9)$$

Each weight has an initial value of 0.5 at the beginning of the training. The weights are dynamically modified after each epoch throughout the training based on their relative contribution to the validation loss. This method optimizes the model to minimize loss functions based on their relative importance in the overall validation loss, improving overall performance and achieving a better balance between reconstruction accuracy and perceptual quality.

3.3.1 Gradient Magnitude Similarity Deviation (GMSD)

To measure the similarity between two images, the Low-Resolution Image (LRI) (a distorted image) and the High-Resolution Image (HRI) (a ground truth image), W. Xue et al. created the Gradient Magnitude Similarity Deviation (GMSD) in 2014. GMSD evaluates how the global variation of the gradient-based local-quality map is used. GMSD is highly consistent with how people perceive the quality of an image and is computationally efficient. GMSD is robust to various visual artifacts, such as noise, blur and compression [31].

GMSD is calculated using the following steps:

- a. Gradient Magnitude Calculation: Use the Prewitt filter to compute horizontal and vertical gradient magnitudes for the Low-Resolution Image (LRI) and the High-Resolution Image (HRI).
- b. Gradient Magnitude Similarity (GMS): Calculate the Gradient Magnitude Similarity (GMS) between LRI and HRI by considering the relationship between their gradient magnitudes.
- c. Mean GMS: Compute the mean GMS values across the entire image for LRI and HRI.
- d. GMSD Calculation: Determine the GMSD, representing the deviation in gradient magnitude similarity, by assessing the differences between individual GMS values and their mean values.

3.3.2 Charbonnier Loss

The Charbonnier loss is a smooth approximation of the Huber loss that preserves its robustness while being more differentiable and simpler to optimize. Huber loss has a non-smooth quadratic to linear transition, which makes gradient-based optimization challenging. The epsilon parameter in the Charbonnier loss regulates the smoothness of the transition from the quadratic to linear region. The Charbonnier loss acts like the L2 loss (MSE) when the epsilon is small and the L1 loss (MAE) when the epsilon is large. For this experiment, the epsilon was set to 0.001. For image-restoration tasks, like denoising, deblurring and super-resolution, Charbonnier loss is better suited, because it penalizes substantial errors less severely than the MSE loss. Compared to MSE loss, Charbonnier-loss outcomes are more visually pleasing. MSE loss function may produce fuzzy edges. This occurs, because the Maximum Likelihood Estimator (MLE) for MSE is the arithmetic mean, while the edges in images typically have two unique modes or values, making them bimodal. Charbonnier loss is expressed by Equation (10).

$$L(y, y') = \frac{1}{n} \sum \sqrt{(y - y')^2} + \epsilon^2 \quad (10)$$

where, y and y' are the original and predicated images, respectively, while ϵ is a small constant.

For minor errors, the square root term in the formula behaves like an L1-norm penalty; for more enormous errors, it behaves like an L2-norm penalty.

3.3.3 Optimizer

"Reduce LR on Plateau" technique at run time dynamically decreases the learning rate by a predetermined factor when validation loss of the model stops improving after a certain number of iterations. For this study, initial learning rate was set to $1e-2$ and learning rate reduction factor was 10% with a waiting of 3 epochs. At around 73% accuracy, the model started to show prolonged improvement, after which the learning rate was changed to $1e-3$; finally, at 98.16% accuracy, it was changed to $1e-4$.

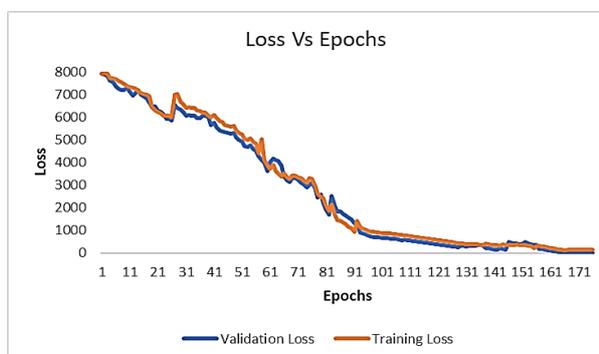
3.4 Dataset Preparation

The dataset used for this research is derived from the CURE-TSR dataset [1]. The CURE-TSR dataset contains over two million images of traffic-sign symbols cropped from the CURE-TSD video dataset. The CURE-TSR dataset is introduced to analyze and evaluate the efficiency of algorithms under challenging conditions. The CURE TSR dataset contains 14 different sign types and 12 challenging conditions. For this study, we created two subsets of the CURE TSR dataset. The first subset is the real-image dataset (RID) and the second is the synthesized-image dataset (SID). The real-image dataset (RID) contains 84K images with five challenging conditions and 14 sign types. The SID dataset contains 16800 images and is mainly used to evaluate the model and compare the results with the RID test dataset.

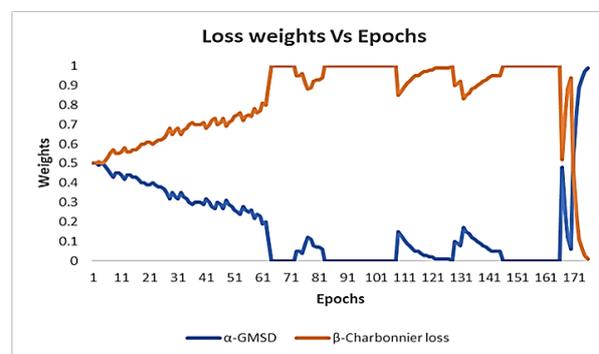
The RID dataset is partitioned into Training RID and Testing RID sub-sets with a ratio of 80:20. The training dataset contains 67200 images and the testing dataset contains 16800 images. Similarly, the test SID dataset contains 16800 synthesized images. This study considered 14 traffic signs with five visual challenges: codec error, dirty lens, rain, shadow and haze images.

4. TRAINING AND VALIDATION LOSS

Graph 1 shows the training *versus* validation loss for the proposed model developed to restore traffic-sign images. The x-axis signifies the number of epochs (i.e., the number of times the model has been trained on the entire dataset). At the same time, the y-axis represents the loss, which measures the distinction between the predicted output of the model and the actual output. The training loss and validation loss were relatively high during the initial training process, indicating that the model was not accurately predicting the restored images. However, as the number of epochs increased, the training loss and validation loss decreased gradually. This suggests that the model was learning to restore the images better over time.



Graph 1. Training loss and validation loss.



Graph 2. Values of alpha and beta.

Initially, the training loss was 7935 and the validation loss was 7922. After 178 epochs, the training loss and validation loss dropped to 186 and 139, respectively, which shows that the model has learned the weights much more accurately. Further, there was no improvement in the loss. This encouraging outcome indicates that the model can restore traffic-sign images with high accuracy and could be used for practical applications.

4.1 Loss Tuning Parameters-Alpha and Beta

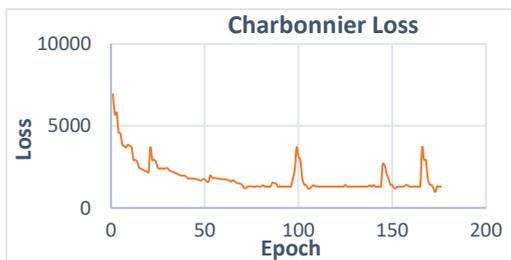
As shown in Equation (9), the alpha and beta weights are the weights of the GMSD and Charbonnier losses, respectively. Updating these weights gives more weight to the more critical loss function and less weight to the less important loss function. Suppose that the ratio of the validation loss to the training loss decreases. In that case, the model performs better on the validation sub-set than on the training sub-set, which may indicate that the Charbonnier loss contributes more to the total loss. In this case, the beta weight is increased and the alpha weight is decreased; so, the model gives more weight to the Charbonnier loss.

On the other hand, if the ratio of the validation loss to the training loss increases, this indicates that the model performs worse on the validation sub-set than on the training sub-set. This could happen if the model overfitted to the training sub-set, which means that it fits the noise in the training sub-set rather than the underlying pattern. In such cases, the GMSD loss, which measures the structural similarity between the ground truth and the predicted image, becomes more critical, as it encourages the model to generate visually similar images to the ground truth rather than just fitting the training-set noise. Therefore, increasing the alpha weight and decreasing the beta weight ensure that the model gives more weight to the GMSD loss and tries to generate visually similar images to the ground truth rather than just fitting the training-set noise.

The decay parameter controls the rate at which the weights are updated, which increases or decreases the weights in each epoch. The decay parameter is a ratio of validation loss to training loss. A small constant (epsilon) is added to prevent alpha and beta from becoming absolute zero. If either alpha or beta becomes zero, the corresponding loss function will not contribute to the overall loss and the model will not be able to learn any further from that loss function. The alpha and beta factors are clamped between a minimum value (epsilon) and a maximum value (1.0). The sum of alpha and beta is fixed at 1.0; therefore, the distribution graph of alpha and beta looks like a mirror image of each other, as shown in Graph 2.

4.2 Comparing Hybrid Loss Function and Single Loss Functions

As discussed in sub-section 3.2, to demonstrate the effectiveness of our novel custom loss function that combines the GMSD (Gradient Magnitude Similarity Deviation) and Charbonnier loss functions, we conducted ablation experiments, as shown in Graphs 3-6; the proposed hybrid loss has optimal convergence of 139 after 178 epochs. Charbonnier, GMSD and MSE have optimal losses of 1206, 1238 and 0.211, respectively. The proposed hybrid loss also has smooth progression compared to other individual loss functions.



Graph 3. Charbonnier loss.



Graph 4. GMSD loss.



Graph 5. Hybrid loss (Proposed).



Graph 6. MSE loss.

During the experiment, we also captured quantitative matrices PSNR, SSIM and MAE of the model with each of the loss functions; as shown in Table 1. The proposed hybrid function exceeds all the parameters except for RMSE, where MSE loss has the least value of 105.651 against 106.151.

Table 1. Loss-function comparison.

Loss Function	PSNR	SSIM	MAE
Hybrid	41.024	0.9726	0.040
MSE	31.940	0.9169	0.090
Charbonnier	29.091	0.8628	0.127
GMSD	28.315	0.8390	0.164

5. MODEL PERFORMANCE EVALUATION

The model's performance has been evaluated based on the image-processing domain's three most frequent evaluation metrics. These are: The Mean Absolute Error (MAE), Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM).

5.1 Mean Absolute Error (MAE)

MAE measures the average absolute difference between the restored and ground truth images. Lower MAE indicates a better resemblance of the input image with the ground truth image. Equation (11) is used to calculate MAE.

$$MAE = \frac{1}{N} * \sum |y_i - x_i| \quad (11)$$

where, N represents the total number of pixels in the restored and ground truth images, y_i and x_i are the pixel values at the corresponding positions in the restored and ground truth images. The summation is taken over all pixels in the images. The absolute value (|.|) ensures that the differences are positive and the average is calculated by dividing the sum by N. Table 2 shows the MAE values when the model is tested on the total test images in individual challenging conditions.

Table 2. MAE for each challenging condition in the test dataset.

Challenging Condition	No. of Signs	Number of Images	MAE value for Test RID			MAE value for Test SID		
			Min.	Max.	Avg.	Min.	Max.	Avg.
Codec Error	14	2800	0.0140	0.046	0.022	0.0146	0.048	0.022
Dirty Lens	14	2800	0.0026	0.038	0.018	0.0028	0.044	0.018
Rainy Images	14	2800	0.0034	0.042	0.020	0.0036	0.042	0.020
Shadow Images	14	2800	0.0012	0.032	0.016	0.0014	0.036	0.018
Haze	14	2800	0.0038	0.042	0.026	0.0038	0.042	0.026

5.2 Peak Signal-to-Noise Ratio (PSNR)

PSNR calculates the relationship between the signal's highest possible power and the noise's power, which affects its accuracy in representing the signal. The equation that defines PSNR is Equation (12).

$$PSNR = 10 \log \left(\frac{R^2}{MSE} \right) \quad (12)$$

where, R is the maximum pixel value of the image, which is 255 for a coloured image. MSE is the mean squared error between the original image and the restored image. Equation (13) describes the calculation of MSE between the y-original image and y' – restored image for n-data points.

$$MSE = \frac{1}{n} \sum_1^n (y - y')^2 \quad (13)$$

Table 3 shows the PSNR values when the model is tested on the total test images in individual challenging conditions.

Table 3. PSNR for each challenging condition in the test dataset.

Challenging Condition	No. of Signs	Number of Images	PSNR value for Test RID			PSNR value for Test SID		
			Min.	Max.	Avg.	Min.	Max.	Avg.
Codec Error	14	2800	36.45	39.89	38.17	34.56	40.02	37.86
Dirty Lens	14	2800	38.23	41.57	39.90	37.95	40.86	38.54
Rainy Images	14	2800	39.78	42.58	41.18	39.89	43.53	39.44
Shadow Images	14	2800	42.68	44.34	43.52	42.34	45.46	43.34
Haze	14	2800	32.84	36.74	34.79	31.08	37.98	34.86

5.3 Structural Similarity Index Measure (SSIM)

SSIM is a metric used to measure the similarity between two images. SSIM is commonly used to evaluate how well a restored image matches the original image. The mathematical formula for SSIM is given in Equation (14).

$$SSIM(x,y) = \frac{(2 * \mu_x * \mu_y + C1) * (2 * \sigma_{xy} + C2)}{(\mu_x^2 + \mu_y^2 + C1) * (\sigma_x^2 + \sigma_y^2 + C2)} \quad (14)$$

where, x and y are the two images being compared, μ represents the mean of an image, σ represents the standard deviation and σ_{xy} represents the covariance of the two images. C1 and C2 are constants that stabilize the division by the weak denominator. Table 4 shows the SSIM values when the model is tested on test images in individual challenging conditions of both test datasets.

Table 4. SSIM for each challenging condition in the test dataset.

Challenging Condition	No. of Signs	Number of Images	SSIM value for Test RID			SSIM value for Test SID		
			Min.	Max.	Avg.	Min.	Max.	Avg.
Codec Error	14	2800	0.716	0.980	0.977	0.782	0.979	0.963
Dirty Lens	14	2800	0.834	0.982	0.979	0.824	0.972	0.968
Rainy Images	14	2800	0.886	0.987	0.976	0.842	0.982	0.976
Shadow Images	14	2800	0.898	0.988	0.982	0.912	0.984	0.978
Haze	14	2800	0.742	0.926	0.934	0.788	0.965	0.924

Overall, the SSIM values suggest that both datasets contain restored images that are highly similar to the original images, with some variations in the levels of similarity across different restored images in each dataset. Thus, the proposed method provides better image restoration even if the traffic sign is captured in challenging conditions.

5.4 Performance Enhancement with Self-attention and Variance-guided Multiscale Attention

An ablation experiment has been performed to evaluate the contribution of self-attention and VGMA in the proposed architecture. Results have been gathered using a real image dataset (RID), as shown in Table 5. Initially, the model was built without any attention mechanism where average PSNR and SSIM were obtained as 30.35 and 0.4525, respectively. In the next iteration, we added only a self-attention module and achieved an average PSNR of 34.46 and an average SSIM of 0.6642. Later on, the model was created using only VGMA and got an average PSNR of 38.42 and an average SSIM of 0.8288. Finally, both self-attention and VGMA modules were added to the model to obtain an average PSNR of 41.024 and an average SSIM of 0.9726. This experiment demonstrates that the proposed model performs better after including both attention modules.

Table 5. Contribution of self-attention and VGMA.

Challenge Type	No attention		Self-attention		VGMA		Self + VGMA	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Codec Error	26.36	0.1235	29.53	0.8109	34.47	0.8094	39.89	0.9800
Dirty Lens	31.65	0.2833	38.98	0.4724	42.45	0.8012	41.57	0.9820
Rain	34.58	0.6285	37.87	0.5572	39.71	0.7243	42.58	0.9870
Shadow	28.88	0.5300	38.76	0.8613	42.94	0.8918	44.34	0.9880
Haze	30.28	0.6970	27.16	0.6191	32.55	0.9175	36.74	0.926
Average	30.35	0.4525	34.46	0.6642	38.42	0.8288	41.024	0.9726

5.5 Experimental Results

Figures 4-8 show the model results for a few test images in each challenging condition: Codec Error, Dirty Lens, Rain, Shadow and Haze. As shown in the figure, the input image in the first column(a) for every table is effectively restored by the model image shown in column (b), which is almost similar to the ground truth image shown in column (c).

Figure 4. Codec Error.		Figure 5. Dirty Lens.		Figure 6. Rainy Images.		Ground Truth Images (c)
I/P Image (a)	Restored Image (b)	I/P Image (a)	Restored Image (b)	I/P Image (a)	Restored Image (b)	

Figure 7. Images with Shadow.		Figure 8. Hazy Images.		Ground Truth Images (c)
I/P Image (a)	Restored Image (b)	I/P Image (a)	Restored Image (b)	



5.6 Comparative Analysis

The comparative analysis of CDRSHNet against existing methods in the literature reveals the model's significant prowess in diverse image-restoration challenges. As shown in Table 6, when addressing codec errors, CDRSHNet achieves a PSNR of 39.89 and an SSIM of 0.9800 for single images, outperforming other models not specifically tailored for this challenge. The model excels for dirty-lens correction, presenting a PSNR of 41.57 and an SSIM of 0.9820, surpassing previous techniques designed for single images and videos. CDRSHNet outshines prominent models in rain restoration, exhibiting a PSNR of 42.58 and an SSIM of 0.987, highlighting its superiority in handling rain-induced distortions. Moreover, in shadow removal, CDRSHNet remarkably outperforms existing techniques, demonstrating a PSNR of 44.34 and an SSIM of 0.9880, showcasing its effectiveness in eliminating shadows. However, in the domain of haze restoration, while CDRSHNet displays competitive results with a PSNR of 36.74 and an SSIM of 0.926, there are no metrics from previous methods in the literature for direct comparison. Overall, the CDRSHNet model consistently showcases superior performance across diverse image-restoration challenges, setting a new benchmark in the field.

Table 6. Comparison with existing methods.

Method	Challenge Type	Image / Video	PSNR	SSIM
Proposed	Codec Error	Single Image	39.89	0.9800
X. Li et al., [14]	Dirty lens	Video	35.37	0.9800
Y. Wang et al., [33]	Dirty lens	Single Image	23.43	0.8640
Proposed	Dirty lens	Single Image	41.57	0.9820
H. Wang et al., [16]	Rain	Single Image	35.31	0.9448
S. Li et al., [17]	Rain	Single Image	33.75	0.9412
D. Ren et al., [32]	Rain	Single Image	33.78	0.977
Proposed	Rain	Single Image	42.58	0.987
Z. Liu et al., [19]	Shadow	Single Image	25.92	0.9090
H. Fan et al., [21]	Shadow	Single Image	25.70	0.9826
X. Hu et al., [22]	Shadow	Single Image	25.07	0.8930
Proposed	Shadow	Single Image	44.34	0.9880
W. Yi et al., [34]	Haze	Single Image	19.1736	0.8864
D. Zhao et al., [26]	Haze	Single Image	16.032	0.626
Proposed	Haze	Single Image	36.74	0.926

To demonstrate generalization of the model, it is tested on the German Traffic Sign Recognition Benchmark (GTSRB) dataset and the Belgium Traffic Sign Dataset, as per the results in Table 7. The proposed model exhibits commendable performance across these datasets. Starting with the CURE-TSR dataset, the model achieves a substantial PSNR of 41.02, indicating high fidelity in image reconstruction. At the same time, the SSIM of 97.26 denotes a robust structural similarity to the original images. On the

GTSRB dataset, PSNR and SSIM show improvement, with values of 41.96 and 98.34, respectively. This signifies enhanced image-restoration capabilities, excelling in preserving image quality and structural details. The model's performance peaks on the BelgiumTS dataset, recording a PSNR of 42.62 and an impressive SSIM of 98.96. These values reflect the model's exceptional ability to restore images, surpassing its performance on previous datasets. In summary, the model consistently delivers superior results across datasets, with higher PSNR and SSIM values indicative of its proficiency in reconstructing images while maintaining structural fidelity, making it exceptionally reliable for scenarios where preserving fine details is paramount.

Table 7. PSNR and SSIM across diverse datasets.

CURE-TSR Dataset	PSNR	41.02
	SSIM	97.26
GTSRB Dataset	PSNR	41.96
	SSIM	98.34
BelgiumTS Dataset	PSNR	42.62
	SSIM	98.96

6. CONCLUSION

In this paper, we presented a novel CDRSHNet architecture for restoring traffic-sign images affected by different types of image degradation, including rain, shadow and haze, images taken with dirty lenses and corrupted due to codec error. This work introduces the fusion of self-attention and variance-guided multiscale attention modules with a custom-made loss function to effectively restore the images captured in adverse conditions. The proposed method uses varying learning-rate techniques, group normalization and dilation for better model performance. The experimental results show that the proposed model effectively restores images with high-quality metrics. The overall SSIM average of 0.978 for the Test RID dataset and an overall SSIM average of 0.963 for the Test SID dataset indicate the high restoration quality of our model. Similarly, the overall average PSNR values of 39.58 and 39.46 for Test RID and Test SID datasets, respectively, with an overall accuracy of 99.3%, further confirm the superior performance of the proposed novel architecture. The proposed model exhibits commendable performance across different datasets.

REFERENCES

- [1] D. Temel, G. Kwon, M. Prabhushankar and G. AlRegib, "CURE-TSR: Challenging Unreal and Real Environments for Traffic Sign Recognition," arXiv (Cornell University), DOI: 10.48550/arxiv.1712.02463, Dec. 2017.
- [2] J. Su, B. Xu and H. Yin, "A Survey of Deep Learning Approaches to Image Restoration," *Neurocomputing*, vol. 487, pp. 46–65, DOI: 10.1016/j.neucom.2022.02.046, May 2022.
- [3] Z. Shen and D. Dang, "Mixed Hierarchy Network for Image Restoration," arXiv (Cornell University), DOI: 10.48550/arxiv.2302.09554, Feb. 2023.
- [4] M. Maru and M. C. Parikh, "Image Restoration Techniques: A Survey," *Int. Journal of Computer Applications*, vol. 160, no. 6, pp. 15–19, DOI: 10.5120/ijca2017913060, Feb. 2017.
- [5] L.-Y. Chang and A. I. Kirkland, "Comparisons of Linear and Nonlinear Image Restoration," *Microscopy and Microanalysis*, vol. 12, no. 6, pp. 469–475, DOI: 10.1017/s1431927606060582, Oct. 2006.
- [6] Z. Liu, "Literature Review on Image Restoration," *Journal of Physics, Conference Series*, vol. 2386, no. 1, p. 012041, IOP Publishing, DOI: 10.1088/1742-6596/2386/1/012041, Dec 2022.
- [7] L. Yu, J. Guo and Y. Chen, "Research Status and Development Trend of Image Restoration Technology," *Journal of Physics*, vol. 2303, no. 1, DOI: 10.1088/1742-6596/2303/1/012081, 2022.
- [8] C. Zhang, F. Du and Y. Zhang, "A Brief Review of Image Restoration Techniques Based on Generative Adversarial Models," *Lecture Notes in Electrical Engineering*, pp. 169–175, DOI: 10.1007/978-981-32-9244-4_24, 2019.
- [9] S. Ahmed, U. Kamal and Md. K. Hasan, "DFR-TSD: A Deep Learning Based Framework for Robust Traffic Sign Detection under Challenging Weather Conditions," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–13, DOI: 10.1109/tits.2020.3048878, 2021.
- [10] R. Huang, Y. Zhang and Z. Luo, "Inpainting of Compressed Images with Autoencoder-based Prior Learning," *Proc. of the 26th ACM Int. Conf. on Multimedia*, 236-244.

- [11] S. Jeon, H. Kim and H. Kwon, "Compressed Image Restoration Using Autoencoder Regularization," *Journal of Imaging Science and Technology*, vol. 63, no. 6, pp.060403-1 - 060403-11, DOI: 10.2352/J.ImagingSci.Technol.2019.63.6.060403, 2019.
- [12] K. Zhang, Y. Li and Y. Wang, "A Two-stage Method for Video Codec Error Concealment Using Deep Learning," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 7, pp. 2102-2115, DOI: 10.1109/TCSVT.2020.2977168, 2020.
- [13] M. Uricar et al., "Let's Get Dirty: GAN Based Data Augmentation for Camera Lens Soiling Detection in Autonomous Driving," *Proc. of the IEEE/CVF Winter Conf. on Applications of Computer Vision*, pp. 766-775, [Online], Available: <http://arxiv.org/pdf/1912.02249.pdf>, Dec. 2021.
- [14] X. Li, B. Zhang, J. Liao and P. V. Sander, "Let's See Clearly: Contaminant Artifact Removal for Moving Cameras," *Proc. of the Int. Conf. on Computer Vision*, pp. 2011-2020, Montreal, Canada, Oct. 2021.
- [15] J. Mohd, Sandra Mamani Reyes and J. Xiao, "Camera Lens Dust Detection and Dust Removal for Mobile Robots in Dusty Fields," *Proc. of the 2021 IEEE Int. Conf. on Robotics and Biomimetics (ROBIO)*, DOI: 10.1109/robio54168.2021.9739233, Dec. 2021.
- [16] H. Wang, Y. Wu, Q. Xie, Q. Zhao, Y. Liang et al., "Structural Residual Learning for Single Image Rain Removal," *Knowledge-based Systems*, vol. 213, p. 106595, Feb. 2021.
- [17] S. Li, W. Ren, J. Zhang, J. Yu and X. Guo, "Single Image Rain Removal *via* a Deep Decomposition-Composition Network," *Computer Vision and Image Understanding*, vol. 186, pp. 48-57, Sep. 2019.
- [18] M. Umair Arif, M. U. Farooq, R. H. Raza, Z. U. A. Lodhi and M. A. R. Hashmi, "A Comprehensive Review of Vehicle Detection Techniques under Varying Moving Cast Shadow Conditions Using Computer Vision and Deep Learning," *IEEE Access*, vol. 10, pp. 104863-104886, 2022.
- [19] Z. Liu, H. Yin, Y. Mi, M. Pu and S. Wang, "Shadow Removal by a Lightness-guided Network with Training on Unpaired Data," *IEEE Transactions on Image Processing*, vol. 30, pp. 1853-1865, Jan. 2021.
- [20] H. van Le and D. Samaras, "From Shadow Segmentation to Shadow Removal," *arXiv (Cornell University)*, DOI: 10.48550/arxiv.2008.00267, Aug. 2020.
- [21] H. Fan, M. Han and J. Li, "Image Shadow Removal Using End-to-End Deep Convolutional Neural Networks," *Applied Sciences*, vol. 9, no. 5, p. 1009, DOI: 10.3390/app9051009, Mar. 2019.
- [22] X. Hu, Y. Jiang, C.-W. Fu and P.-A. Heng, "Mask-ShadowGAN: Learning to Remove Shadows from Unpaired Data," *Proc. of the 2019 IEEE/CVF Int. Conf. on Computer Vision (ICCV)*, vol. 2019, pp. 2472-2481, Seoul, S. Korea, Jan. 2019.
- [23] L. Ren, Z. Pan, J. Cao, J. Liao and Y. Wang, "Infrared and Visible Image Fusion Based on Weighted Variance Guided Filter and Image Contrast Enhancement," *Infrared Physics & Technology*, vol. 114, p. 103662, DOI: 10.1016/j.infrared.2021.103662, May 2021.
- [24] Q. Yang, C. Zhang, H. Wang, Q. He and L. Huo, "SV-FPN: Small Object Feature Enhancement and Variance-guided RoI Fusion for Feature Pyramid Networks," *Electronics*, vol. 11, no. 13, pp. 2028-2028, DOI: 10.3390/electronics11132028, Jun. 2022.
- [25] X. Yang, "An Overview of the Attention Mechanisms in Computer Vision," *Journal of Physics: Conference Series*, vol. 1693, p. 012173, DOI: 10.1088/1742-6596/1693/1/012173, Dec. 2020.
- [26] D. Zhao, L. Xu, Y. Yan, J. Chen and L.-Y. Duan, "Multi-scale Optimal Fusion Model for Single Image dehazing," *Signal Processing-Image Communication*, vol. 74, pp. 253-265, DOI: 10.1016/j.image.2019.02.004, May 2019.
- [27] W. Yi et al., "Towards Compact Single Image Dehazing *via* Task-related Contrastive Network," *Expert Systems with Applications*, vol. 235, p. 121130, 2024.
- [28] X. Zhu et al., "GAN-based Image Super-resolution with a Novel Quality Loss," *Mathematical Problems in Engineering*, vol. 2020, p. e5217429, DOI: 10.1155/2020/5217429, Feb. 2020.
- [29] B. Wu, H. Duan, Z. Liu and G. Sun, "SRPGAN: Perceptual Generative Adversarial Network for Single Image Super Resolution," *arXiv (Cornell University)*, DOI: 10.48550/arXiv.1712.05927, Dec. 2017.
- [30] B. Vasant Gajera, S. Raj Kapil, D. Ziaei, J. Mangalagiri, E. L. Siegel and D. Chapman, "CT-Scan Denoising Using a Charbonnier Loss Generative Adversarial Network," *IEEE Access*, vol. 9, pp. 84093-84109, DOI: 10.1109/access.2021.3087424, Jun. 2021.
- [31] W. Xue, L. Zhang, X. Mou and A. C. Bovik, "Gradient Magnitude Similarity Deviation: A Highly Efficient Perceptual Image Quality Index," *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 684-695, DOI: 10.1109/tip.2013.2293423, Feb. 2014.
- [32] D. Ren, W. Zuo, Q. Hu, P. Zhu and D. Meng, "Progressive Image Deraining Networks: A Better and Simpler Baseline," *Proc. of the 2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 3937-3946, DOI: 10.1109/cvpr.2019.00406, Long Beach, USA, Jun. 2019.
- [33] Y. Wang, R. Wan, W. Yang, B. Wen, L.-P. Chau and A. C. Kot, "Removing Image Artifacts from Scratched Lens Protectors," *arXiv (Cornell University)*, DOI: 10.48550/arxiv.2302.05746, Feb. 2023.
- [34] W. Yi, M. Liu, L. Dong, Y. Zhao, X. Liu and M. Hui, "Restoration of Haze-free Images Using Generative Adversarial Network," *Proceedings of the SPIE*, vol. 11432, DOI: 10.1117/12.2541893, Feb. 2020.

ملخص البحث:

في أحوال الطقس الصّعبة، تؤثر أمور متنوّعة على نوعيّة الصّور، ومنها قطرات المطر، والضّلال، والعيوب الناتجة عن عدسات كاميرات مئسّخة، ... وغيرها. وتكافح طرق معالجة هذه العيوب القائمة من أجل التغلّب عليها على نحوٍ شامل؛ إذ إنّ ذلك يتطلّب تقنيات إبداعية وحلولاً مبتكرة لاستعادة الصّور المشوّهة بأيّ من العيوب سالفة الذّكر بحيث تكون أقرب ما يمكن للصّور الأصليّة الخالية من العيوب.

تقدّم هذه الورقة نظاماً هجيناً يستفيد من عددٍ من التقنيات المتنوّعة التي تركّز على معالجة مواقع العيوب في الصّور الملتقطة تحت ظروف الطقس السيئة. وقد تمّ تدريب النّظام المقترح وتقييمه باستخدام مجموعات بيانات ملائمة؛ إذ بلغت دقّة النّظام في استعادة الصّور (99.3%).

وقد جرت مقارنة النّظام المقترح في هذه الدراسة مع عددٍ من الأنظمة الواردة في أدبيات الموضوع، حيث تبين أنّ النّظام المقترح يتفوّق على ما سواه من حيث وضوح الصّور المعالجة وقربها من الصّور الأصليّة.

USING RESNET18 IN A DEEP-LEARNING FRAMEWORK AND ASSESSING THE EFFECTS OF ADAPTIVE LEARNING RATES IN THE IDENTIFICATION OF MALIGNANT BREAST MASSES IN MAMMOGRAMS

Soumia Benbakreti¹, Samir Benbakreti², Kadda Benyahia³ and Mohamed Benouis⁴

(Received: 12-Nov.-2023, Revised: 8-Jan.-2024, Accepted: 25-Jan.-2024)

ABSTRACT

Breast cancer is a prevalent disease that primarily affects women globally, but it can also affect men. Early detection is crucial for better treatment outcomes and mammography is a common screening method. Recommendations for mammograms vary by age and country. Early breast-cancer screening is vital for timely interventions. This paper aims to introduce artificial-intelligence methods through deep-learning approaches utilizing pre-trained CNN-based models for the diagnosis of masses depicted in breast images. These masses may be either malignant or benign, necessitating distinct management strategies for each scenario. The experiments conducted on pre-trained models (AlexNet, InceptionV3 and ResNet18) are designed to underscore the significance of selecting the batch size and adaptive learning rate in influencing the results, ultimately facilitating a notable enhancement in classification rates. Pre-trained models applied to a merged dataset comprising three datasets (Inbreast+MIAS+DDSM) yielded an accuracy of 93.7% for InceptionV3 and 88.9% for AlexNet. However, the most favorable outcome was observed with ResNet18, achieving an accuracy of 95% (with precision, recall and F1-score of 94.90%, 94.91% and 94.91%, respectively).

KEYWORDS

Malignant breast cancer, Lesion classification, Transfer learning, Residual network (ResNet18), Adaptive learning rate.

1. INTRODUCTION

Breast cancer remains a significant concern in global public health, impacting millions of women annually and profoundly affecting their well-being. In 2018, the International Agency for Research on Cancer documented a global 18,078,957 cancer cases. Among these cases, breast cancer constituted 2,088,849 instances, accounting for 11.6% of all cancers and ranking as the second most prevalent type. Notably, breast cancer emerged as the most frequently diagnosed cancer among women, making up 24.2% of the cases [1]. Despite substantial progress in medical research, breast cancer poses ongoing complexities in terms of early detection, effective treatment and long-term management. This multifaceted disease exhibits considerable variations in clinical presentations, disease progression and responses to therapies [2]. Consequently, gaining a deeper understanding of its underlying mechanisms, risk factors and innovative diagnostic and treatment approaches is essential for enhancing the clinical outcomes of breast-cancer patients. CNNs have played a crucial role in enhancing breast-cancer diagnosis [3][4][5]. CNN models, such as ResNet and Inception, have been adapted for the analysis of mammographic images, enabling more accurate detection of suspicious masses and micro-calcifications. There are several methods for diagnosing breast cancer, including mammography [6], which is commonly used in women over 40 clinical-breast examination [7] performed by a healthcare professional to detect anomalies, breast ultrasound that uses sound waves to visualize breast tissues [8] and breast MRI, which is more sensitive than mammography and often used in high-risk women [9]. Biopsy, involving the collection of a tissue sample for laboratory analysis, remains the most precise method for diagnosing breast cancer and determining its type [10]. Additionally, a sentinel lymph node biopsy may be necessary to assess the spread of cancer to lymph nodes. For women with a family history of breast cancer, genetic tests, such as BRCA1 and BRCA2 mutations, can be performed [11]. Finally, in some cases, advanced imaging tests, like CT scans or bone scans, are needed to evaluate the extent

1. S. Benbakreti is with Laboratory of Mathematic, University of Djillali Liabes, Sidi Bel Abbas, Algeria. Email: souben2223@gmail.com
2. S. Benbakreti is with ENSTTIC, Department of Speciality, Street of Senia, Oran, 31000, Algeria. Email: samir.benbakreti@ensttic.dz
3. K Benyahia is with LTC Laboratory, University of Tahar Moulay, Saida, Algéria. Email: benyahiak@gmail.com
4. M Benouis is with University of Augsburg, Augsburg, Germany. Email: mhbenouis@gmail.com

of cancer [12]. The choice of method depends on individual factors and should be discussed with a healthcare professional. On the other hand, the classification of breast tumors into malignant (cancerous) and benign (non-cancerous) categories is a critical task in both medical imaging and oncology. This classification presents specific challenges, notably the visual similarity between certain malignant and benign tumors in medical images (see Figure 1), which complicates their differentiation.

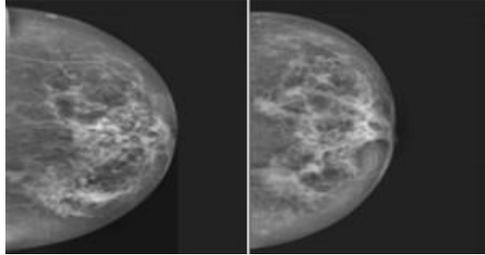


Figure 1. a- Benign, b- Malignant.

Ultimately, our aim is to contribute to the improvement of patient care for those grappling with breast cancer by showcasing the latest scientific and clinical breakthroughs, while also underscoring the persistent challenges warranting specific focus within the medical and research communities. This research focused on constructing a convolutional neural network (CNN) model to differentiate between benign and malignant breast-cancer tumors. Notably, the study emphasizes three key contributions. Firstly, the careful selection of the batch size is highlighted for its substantial impact on image-classification tasks within a CNN. Secondly, the adoption of an adaptive learning rate is advocated to address issues, such as model stability and overfitting. Lastly, the integration of three reference datasets (Inbreast + DDSM + MIAS) is underscored, enhancing the model's adaptability to diverse dataset characteristics and resulting in a more robust and efficient classification model.

The structure of the present document is as follows: An introduction is presented in Section 1 and research works on the early identification of breast cancer are discussed in Section 2. The proposed model and dataset are described in Sections 3 and 4. Section 5 details the experiments performed on the different architectures and the conclusions that were reached.

2. RELATED WORKS

Artificial intelligence (AI) is transforming the medical landscape through its provision of advanced diagnostic solutions. In brain-cancer early detection [26], AI systems meticulously analyse medical images to identify initial signs, thereby enhancing the chances of successful treatment. In neurological diagnostics, AI adeptly interprets brain images, simplifying the detection of diseases, like Alzheimer's [37]. Additionally, AI algorithms analyse ECG data for cardiac assessment [38], enabling the diagnosis of cardiac issues and facilitating preventive interventions. In [39], the authors employ deep-learning algorithms to analyse fundus images, distinguishing early signs and allowing differentiation between multiple ocular diseases. In [42], a system integrating advanced deep-learning networks (EfficientNet, Xception, MobileNetV2, InceptionV3 and Resnet50) through the innovative method of adaptive consensus weighting (CAW) was proposed. The method enables dynamic adjustment of multiple deep networks, thereby enhancing the system's detection capabilities. Evaluations on various datasets, including DDSM, demonstrated quite interesting performances. On the other hand, the study of [43] explored the potential of artificial intelligence (AI) and machine learning (ML) in providing precise diagnoses, enhancing outcome predictions and recognizing disparities in the treatment of CLTI (Chronic Limb-Threatening Ischemia). The authors underscored the importance of AI/ML approaches in patient management and highlighted how existing data could be leveraged for computer-guided interventions. These applications underscore the growing significance of AI in medical diagnosis, paving the way for swifter and personalized treatments. Subsequently, we will delve into advancements related to breast cancer in this context.

In [13], the most recent machine learning-based models for detecting and classifying breast cancer are examined through a comparative study. Additionally, a compilation of widely accessible and well-received datasets is offered to facilitate future experiments and comparisons. The results of the analysis

highlight You Only Look Once (YOLO) and RetinaNet as the most accurate recent models for both detection and classification. But, the YOLO model can be tailored to specific requirements; however, it does exhibit limitations when it comes to handling closely positioned objects. In particular, it struggles to achieve high accuracy with small-sized objects and may occasionally make errors in localization. Similarly, RetinaNet has certain shortcomings, like model overloading and complexity, which make training so difficult. The study of Ismail & Sovuthy made a comparative analysis of breast-cancer detection using two deep-learning model networks. The entire process encompasses image preprocessing, classification and performance assessment [14]. Specifically, the performance of two deep-learning model networks, VGG16 and ResNet50, is assessed for distinguishing between normal and abnormal tumors utilizing the IRMA dataset. The findings indicate that VGG16 outperforms ResNet50 in terms of accuracy, achieving a 94% accuracy rate compared to ResNet50's 91.7%. Although the study was conducted on a database containing a very limited number of samples (1515 images), the results obtained are excellent, but cannot be generalized. In article [15], the authors aim to conduct a comprehensive examination of machine-learning techniques and their practical applications in diagnosing and predicting outcomes for breast cancer in the context of British Columbia. They begin by offering an overview of various ML techniques, including artificial neural networks (ANNs), support vector machines (SVMs), decision trees (DTs) and k-nearest neighbors (k-NNs). Subsequently, they apply these techniques specifically to the breast cancer scenario in British Columbia, utilizing primary data sourced from the Wisconsin Breast Cancer Database (WBCD) as a benchmark for comparing results obtained through different algorithms. Lastly, they also introduce a healthcare-system model derived from their recent-research endeavors. However, traditional-machine learning techniques sometimes need human-feature extraction and have difficulties in processing complex, high-dimensional data. In the provided framework, features are extracted from images using pre-trained CNN architectures, specifically GoogLeNet, Visual Geometry Group Network (VGGNet) and Residual Networks (ResNet). These extracted features are then input into a fully connected layer for classifying malignant and benign cells, employing average pooling for classification. To assess the effectiveness of this framework, experiments were conducted on well-established benchmark datasets. The results demonstrated that the proposed framework achieves an impressive accuracy rate of 97.52% [16]. However, we will note that the framework requires a lot of hardware resources, which are sometimes lacking. The study of [17] introduces an all convolutional-network method for categorizing screening mammograms, surpassing previous approaches. When tested on digitized film mammograms (CBIS-DDSM), the best model achieved an AUC of 0.88 and combining four models improved the AUC to 0.91. Similarly, on a distinct set of full-field digital mammography images (INbreast database), the top model achieved an AUC of 0.95 and averaging results from four models raised the AUC to 0.98.

Furthermore, the research demonstrates that a classifier trained using their method on CBIS-DDSM mammograms can be successfully adapted to INbreast FFDM images with minimal additional data for fine-tuning. These findings highlight the potential of deep-learning methods in enhancing clinical tools and reducing errors in mammography screening. Wang et al. proposed a strategy that utilizes SqueezeNet with fire modules and a complex bypass to extract informative features from mammography images [34]. Subsequently, these extracted features are employed to train a Support Vector Machine (SVM) for the classification of mammography images. The model, known as SNSVM, which integrates SqueezeNet guidance with SVM, exhibited promising results when tested on the MIAS dataset, achieving an accuracy of 94.10%. In [35], the authors devised a novel framework for breast-cancer diagnosis, employing entropy-controlled deep learning and flower-pollination optimization based on mammogram images. Within this proposed framework, a contrast-enhancement method is developed using filter fusion. The pre-trained ResNet50 model undergoes refinement and training through transfer learning on both the original and enhanced datasets, employing various data augmentation techniques. While the results obtained from individual datasets appear highly promising, it's worth noting that the pretrained model, due to its depth, can be resource-intensive during processing. Similar research has been conducted on ResNet101 [36], incorporating feature fusion through the proposed highly corrected function-controlled canonical-correlation analysis approach and optimal feature selection using the Newton-Raphson algorithm controlled by Satin Bowerbird Optimization. The experiments within the devised framework were carried out utilizing the CBIS-DDSM dataset, yielding a top accuracy of 94.5%. Han et al. introduced a novel deep-learning model for the multi-classification of breast cancer. The structured deep learning model has delivered

impressive results, achieving an average accuracy rate of 93.2% on a substantial dataset [18]. Nevertheless, we believe that the exploitation of this deep-structured model can be improved, as well as the results obtained.

3. PROPOSED MODEL

In this research, we will investigate three distinct approaches for detecting masses within breast images. These pre-trained neural network architectures, specifically drawn from CNN [19], include AlexNet, ResNet18 and InceptionV3. We selected these three models considering that resource consumption varies among pre-trained models based on their complexity and parameter count. Typically, deeper and more complex models, like ResNet152 [40] or advanced Inception variants [41], may demand more resources in terms of GPU memory and computing power for training and inference. Thus, our choice aimed to find a compromise between performance and resource consumption. In this study, transfer learning is implemented with deep-learning models to detect suspicious masses in mammograms. The deep-learning model is characterized as a pre-trained model, trained using a Convolutional Neural Network, with features generated through transfer learning. The proposed model is shown in Figure 2.

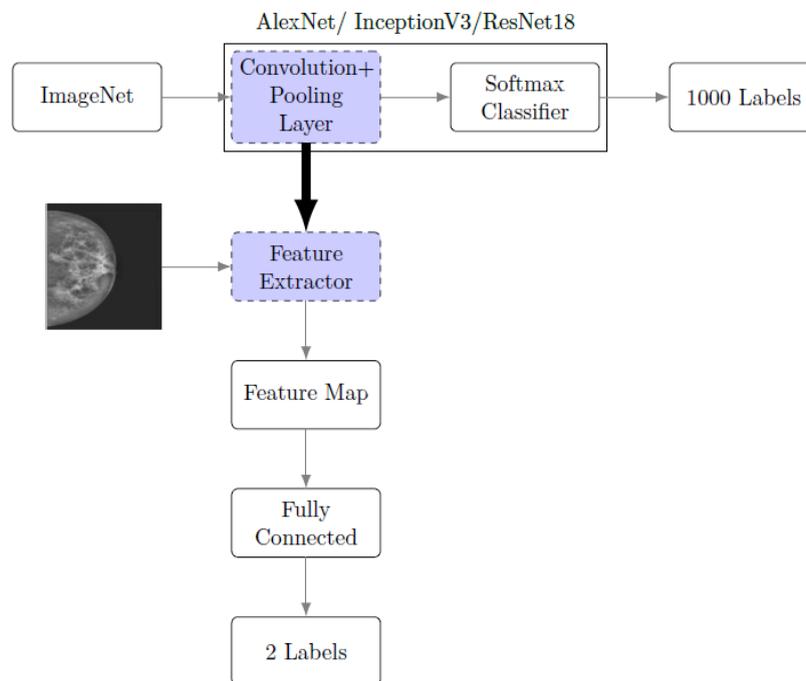


Figure 2. Proposed model.

Utilizing the extracted features, a fully connected layer is employed to construct the final model. The breast-mass identification scheme is illustrated in Figure 2. During the transfer learning and retraining of the deep-learning model, the extracted features are employed to train the new model. The input comprises images and all convolutional and pooling layers are reused in the training of the new model. In essence, ResNet18, InceptionV3 or AlexNet serves as a feature extractor, subsequently transmitted to an FC layer.

3.1 AlexNet

AlexNet is an eight-layer convolutional neural network [20] composed of five convolutional layers and three fully-connected layers. The first two convolutional layers are followed by a normalization layer and a max-pooling layer, while the third and fourth layers are directly connected. The fifth convolutional layer is succeeded by a max-pooling layer. The resulting output undergoes a sequence of two fully-connected layers, where the second fully-connected layer contributes to a Softmax classifier. AlexNet uses ReLU as the activation function, differing from the conventional sigmoid and tanh functions used in previous neural networks. ReLU is a non-saturating activation function that not only significantly accelerates the model's training time, but also more effectively addresses the issues

of gradient disappearance and explosion, making it simpler to train a deeper network. The standard input size for the AlexNet model is $227 \times 227 \times 3$. Table 1 summarizes the AlexNet architecture.

Table 1. The AlexNet architecture.

Model	Alexnet	
	Filter Size/Stride	Output Size
Conv1	11×11/4	96×55×55
Pool1	3×3/2	96×27×27
Conv2	5×5/1	256×27×27
Pool2	3×3/2	256×13×13
Conv3	3×3/1	384×13×13
Conv4	3×3/1	384×13×13
Conv5	3×3/1	256×13×13
Pool5	3×3/2	256×6×6
Fc5	-	4096
Fc6	-	4096
Fc7	-	1000

3.2 InceptionV3

InceptionV3 contains 50 layers and is a convolutional neural network (CNN) [21]. The algorithm, named "going deeper with convolutions" was developed and trained by Google. The development of the Inception module, which consists of a series of 1-by-1 convolutional layers/blocks used for dimensionality reduction and feature aggregation, is the key component of GoogleNet/Inception architecture. This model had 9 inception modules and a total of 22 layers. Up to 1000 objects can be classified using the pre-trained version of InceptionV3 with the ImageNet dataset weights [22]. This network's image input was 299×299 pixels in size.

The InceptionV3 stands out by incorporating a crucial element known as an inception module. This module utilizes receptive kernels of various sizes, maintaining a consistent output size for the convolution operation through the use of zero padding. The ultimate feature maps are obtained by concatenating the filters. The inception operation plays a significant role in extracting more comprehensive features from the input image. Refer to Figure 3 for a visual representation of this module.

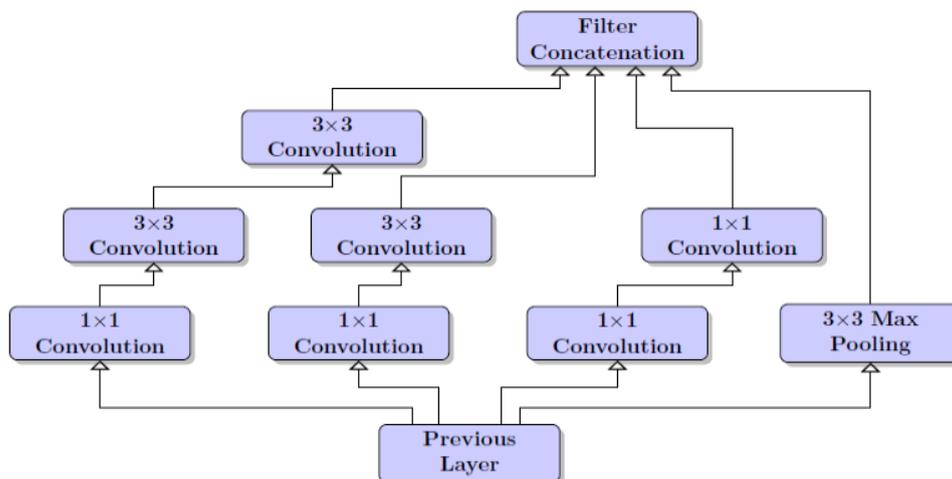


Figure 3. Representation of inception module.

3.3 ResNet18

A convolutional neural network with 18 layers in depth is called ResNet18. Deep Residual Learning for Image Recognition, as it is known, was developed and trained by Microsoft in 2015 ([23]). To address the issue of vanishing gradient that may affect the weightage change in neural networks, ResNet architectures

"Using ResNet18 in a Deep-learning Framework and Assessing the Effects of Adaptive Learning Rates in the Identification of Malignant Breast Masses in Mammograms," S. Benbakreti, S. Benbakreti, K Benyahia and M Benouis.

introduced the use of residual layers and skip connections. This made training easier and allowed neural networks to get much deeper with greater performance. More than a million images from the ImageNet dataset were used to train this model. The network was trained on colored images with a resolution of 224×224 pixels and can categorize up to 1000 objects.

Given its comparatively modest architecture, ResNet18 was selected for its balanced trade-off between depth and performance. It comprises a 7×7 convolutional layer, 2 pooling layers, 5 residual blocks and an FC layer. Each residual block is composed of two 3×3 convolutional layers, followed by a batch-normalization layer and a ReLU activation function (see Table 2).

The network achieves high classification accuracy by employing bottleneck residual blocks, batch normalization for adjusting input layers and identifying connections to mitigate the risk of vanishing gradients.

Figure 4 illustrates the architecture of the ResNet18 network, detailing the configuration of the employed residual blocks (Res Block2 represents a ResNet block with a 1×1 convolution). The notation "FC" denotes a fully-connected layer with two outputs corresponding to malignant and benign classifications. Table 3 gives a summary of the pre-trained models used in this study and their main characteristics.

Table 2. ResNet18 architecture.

Layer Name	Output Size	Description
conv1	$112 \times 112 \times 64$	7×7 , 64, stride 2
conv2_x	$56 \times 56 \times 64$	3×3 max pool, stride 2
		$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$
conv3_x	$28 \times 28 \times 128$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$
conv4_x	$14 \times 14 \times 256$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$
conv5_x	$7 \times 7 \times 512$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$
average pool	$1 \times 1 \times 512$	7×7 average pool
fully connected	1000	512×1000 fully connections
softmax	1000	

Table 3. Summary of pre-trained models.

Pre-trained Model	Depth	Parameters (Millions)	Image Input	Specific feature
AlexNet	8	60	227×227	Deeper
ResNet18	18	11.7	224×224	Residual block
InceptionV3	50	23.9	299×299	Inception module and size kernel

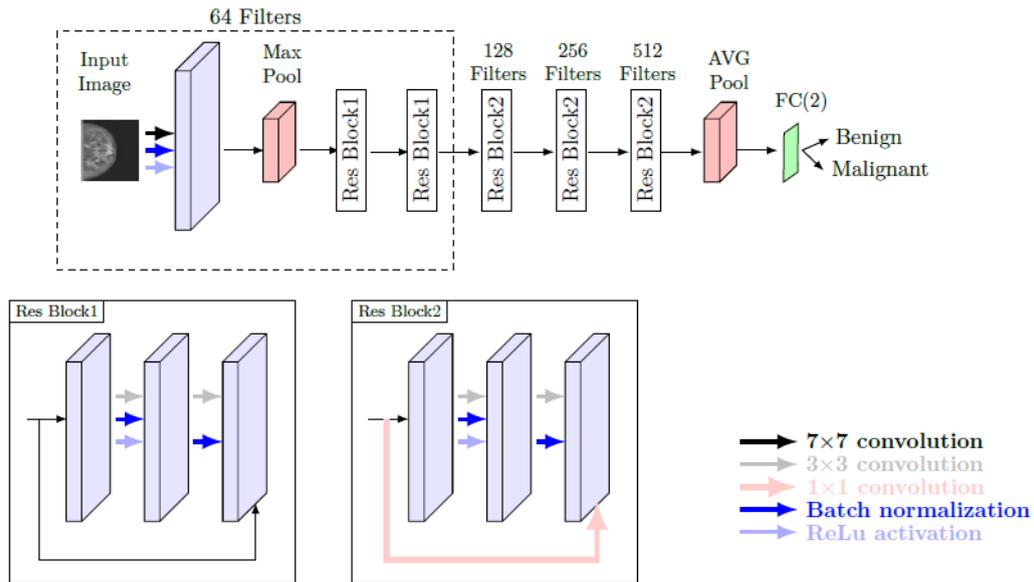


Figure 4. ResNet18 architecture used in the proposed model.

4. THE USED DATASET

The dataset comprises mammograms featuring both benign and malignant masses, resulting from the amalgamation of several datasets [24]. Specifically, it incorporates 7,632 images from the Inbreast dataset, 3,816 images from the MIAS dataset and 13,128 images from the DDSM dataset. All of these images have been uniformly resized to dimensions of 227×227 pixels. The total number of images is 24576. Table 4 displays details about images from the datasets that were used. Matlab 2021 was used for the training. It was installed on a workstation running Windows 10 Pro, 64-bit, with 24 GB of RAM, an Intel(R) Xeon(R) CPU E5-2620 v3 @ 2.40GHz and a NVIDIA Quadro K620 GPU. The majority of the dataset, 80%, is used for training, while 20% is utilized for testing. We refrained from employing data augmentation for two primary reasons.

Firstly, we harnessed the power of transfer learning [25]. Our models were originally trained on extensive datasets encompassing a diverse range of images, such as ImageNet [22]. Consequently, these pretrained models served as our initial framework for the task at hand, which involved classifying breast masses. Transfer learning is especially advantageous when dealing with limited data, as was the case with our dataset, as it allows the model to leverage the knowledge it acquired during previous training. Secondly, we considered existing literature, which suggested that data augmentation does not universally guarantee improved performance, particularly when the introduced transformations lack relevance [26].

Table 4. The used datasets (Inbreast+MIAS+DDSM).

Dataset	Total number of images	Benign	Malignant
Inbreast	7632	2520	5112
MIAS	3816	2376	1440
DDSM	13128	5970	7158
Inbreast+MIAS+DDSM	24576	10866	13710

5. RESULTS AND DISCUSSION

To evaluate the performance of our models, we used the following metrics:

$$Accuracy = \frac{TP+TN}{TP+FN+TN+FP} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

"Using ResNet18 in a Deep-learning Framework and Assessing the Effects of Adaptive Learning Rates in the Identification of Malignant Breast Masses in Mammograms," S. Benbakreti, S. Benbakreti, K Benyahia and M Benouis.

$$F1 - score = \frac{2 * Precision * Recall}{Precision + Recall} \quad (4)$$

Where:

- TP (True Positive): It represents the number of correctly predicted positive samples by the model. This means that the model has correctly identified these samples as positive.
- TN (True Negative): It represents the number of correctly predicted negative samples by the model. This means that the model has correctly identified these samples as negative.
- FP (False Positive): It represents the number of negative samples incorrectly predicted as positive by the model. This means that the model has identified these samples as positive when they are actually negative.
- FN (False Negative): It represents the number of positive samples incorrectly predicted as negative by the model. This means that the model has identified these samples as negative when they are actually positive.

The initial parameters of the three architectures are mentioned in Table 5.

Table 5. The training parameters of the pre-trained models.

Parameter	Value
Initial learning rate	0.001
Optimizer	SGDM
Max. epoch	1
Mini-batch size	20
Activation function	Softmax
Validation frequency	10

When:

- Initial learning rate: It is the initial value of the learning rate in a machine-learning algorithm, especially in neural networks.
- SGDM: the stochastic gradient descent with momentum solver [27].
- Mini-batch size: The stochastic gradient descent technique uses a portion of the training data for each iteration to evaluate the gradient and update the parameters. At each iteration, a distinct subset, known as a mini-batch, is used to assess the gradient of the loss function and update the weights.
- Max. epoch: It is the maximum number of epochs for training.
- Activation function: Transfer functions calculate a layer's output from its net input.
- The validation-frequency value indicates how many iterations there are between evaluations of validation metrics.

The results obtained are shown in Table 6.

Table 6. The obtained results.

Models	The used metrics			
	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
AlexeNet	83.5	85.5	82	83.72
InceptionV3	89.8	90.86	88.87	89.86
ResNet18	88.24	88.13	88	88.07

As depicted in Table 6, it is evident that the three models have produced satisfactory outcomes, with InceptionV3 emerging as the top performer (89.8 % of accuracy). To achieve a more comprehensive and accurate evaluation of the models' performance, we have incorporated supplementary metrics, including

precision, recall and F1-score, alongside accuracy. This approach proves to be particularly valuable in the context of imbalanced class distribution within this scenario.

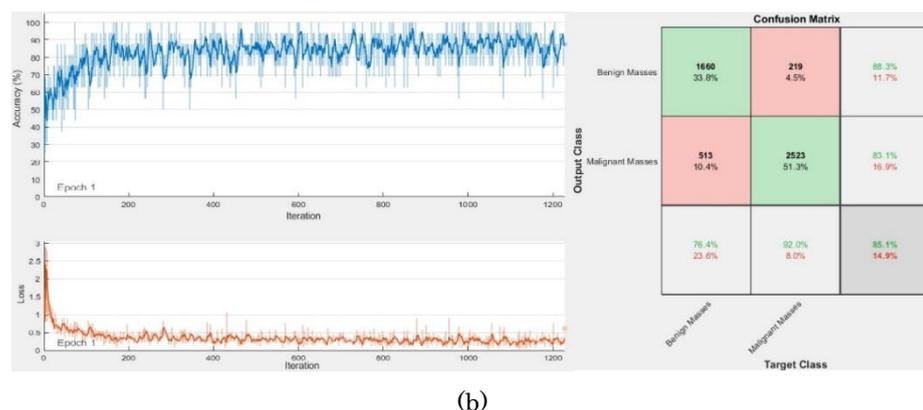
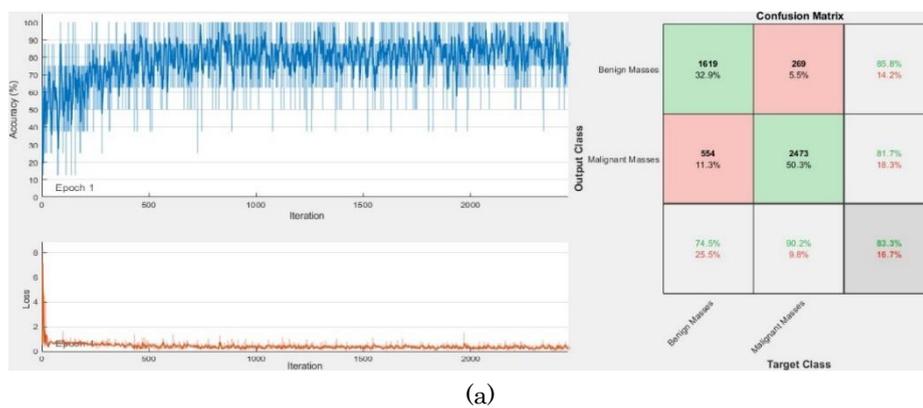
5.1 Experiment 1: Impact of Batch Size on Outcomes

In this experiment, we will explore how altering the batch size influences the results of classification. These results are presented in Table 7.

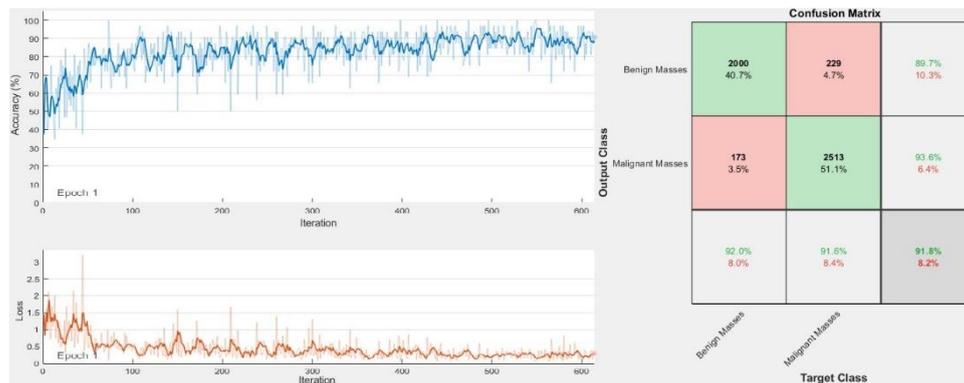
Table 7. The obtained results showing the impact of batch size.

Batch size	Models	The used metrics			
		Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
8	AlexeNet	80.2	83.5	87.2	80.75
	InceptionV3	83.3	83.73	82.35	83.03
	ResNet18	83.1	23.95	83.4	83.17
16	AlexeNet	82.4	82.66	82.73	82.19
	InceptionV3	85.1	85.72	84.20	84.96
	ResNet18	84.8	84.78	84.35	84.56
32	AlexeNet	85.1	84.87	84.84	84.86
	InceptionV3	90.89	90.76	91.21	91.02
	ResNet18	91.8	91.64	91.74	91.74
64	AlexeNet	86	85.94	86.42	86.18
	InceptionV3	88.5	88.64	89.14	88.89
	ResNet18	91.6	91.47	92	91.73

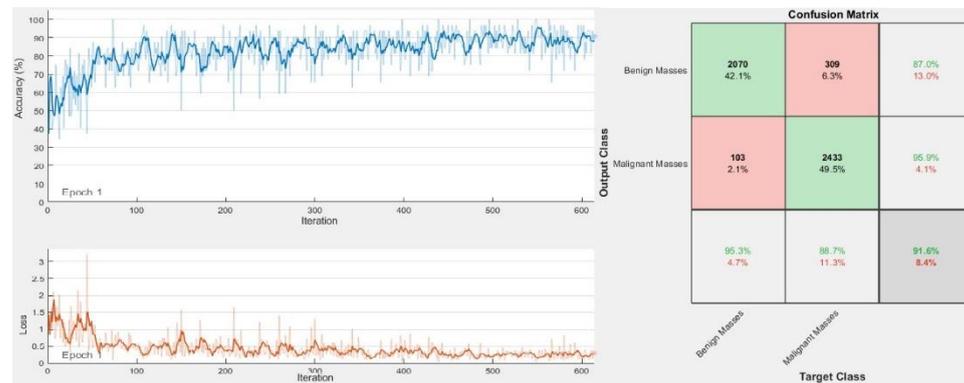
Upon reviewing the curves in Figure 5, we can observe that, apart from the notably quicker convergence seen with a smaller batch size, a larger batch size tends to diminish noise in weight updates (Figure 5(d)). This reduction in noise can enhance training stability and reduce vulnerability to random fluctuations, which proves advantageous, particularly when dealing with noisy datasets or unstable gradients. However, it is important to note that increased stability may not always result in overall improved performance. Conversely, a smaller batch size introduces noise into weight updates, effectively aiding in the prevention of overfitting (Figure 5(a)) [28].



"Using ResNet18 in a Deep-learning Framework and Assessing the Effects of Adaptive Learning Rates in the Identification of Malignant Breast Masses in Mammograms," S. Benbakreti, S. Benbakreti, K Benyahia and M Benouis.



(c)



(d)

Figure 5. (a) Best result for breast-mass classification (InceptionV3) with BS=8. (b) Best result (InceptionV3) with BS=16, (c) Best result (ResNet18) with BS=32, (d) Best result (ResNet18) with BS=64.

Overfitting, a situation where the model overly adapts to the training data and struggles to generalize to validation or test data, is mitigated by the use of a smaller batch size, albeit at the cost of requiring more training iterations. Furthermore, it is worth mentioning that the selection of an appropriate batch size is intricately connected with other model hyper-parameters, such as the learning rate.

This often necessitates an iterative exploration of these hyper-parameters to identify the optimal combination for a specific classification task [29]. The ResNet18 model with a batch size of 32 demonstrated the highest performance in building a transfer-learning model for classifying mammogram-detected breast lesions. Nevertheless, there is not a universally ideal batch size; typically, it necessitates experimentation to determine the best configuration for a specific problem.

5.2 Experiment 2: Impact of Adaptive Learning Rate

In this experimental setup, we maintained the parameters as detailed earlier in Table 5 in particular the same optimizer: SGDM, with the exception of adjusting the batch size to 32 (yielding the best result) and modifying the number of iterations. This modification was made to specifically investigate the influence of the adaptive learning rate. We will utilize a variable learning rate that adapts throughout the training process within a pre-trained model, employing an adaptive learning rate adjustment method. This can enhance the convergence of the learning process.

Table 8. The additional training parameters of the pre-trained models.

Parameter	Value
Max. epoch	4
Learning-rate schedule	Piecewise
Learning-rate drop period	1
Learning-rate drop factor	0.1

In Table 8, the 'LearnRateSchedule' is configured as 'piecewise,' indicating the use of an adaptive learning rate. The 'LearnRateDropPeriod' determines the frequency of learning-rate adjustments, for instance, every 1 epoch and the 'LearnRateDropFactor' specifies the magnitude of reduction at each adjustment, set at 0.1. The classification results of the three pre-trained models are mentioned in Table 9.

Table 9. Adaptive learning-rate results.

Models	The used metrics			
	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
AlexeNet	88.9	88.60	88.73	88.86
InceptionV3	93.7	93.55	93.5	93.57
ResNet18	95	94.90	94.91	94.91

When we compare Tables 7 and 9, it becomes evident that the implementation of adaptive learning rates has notably enhanced the results for all three models. Nevertheless, ResNet18 continues to stand out as the network that achieved the highest classification rate, reaching 95%. (see Figure 6).

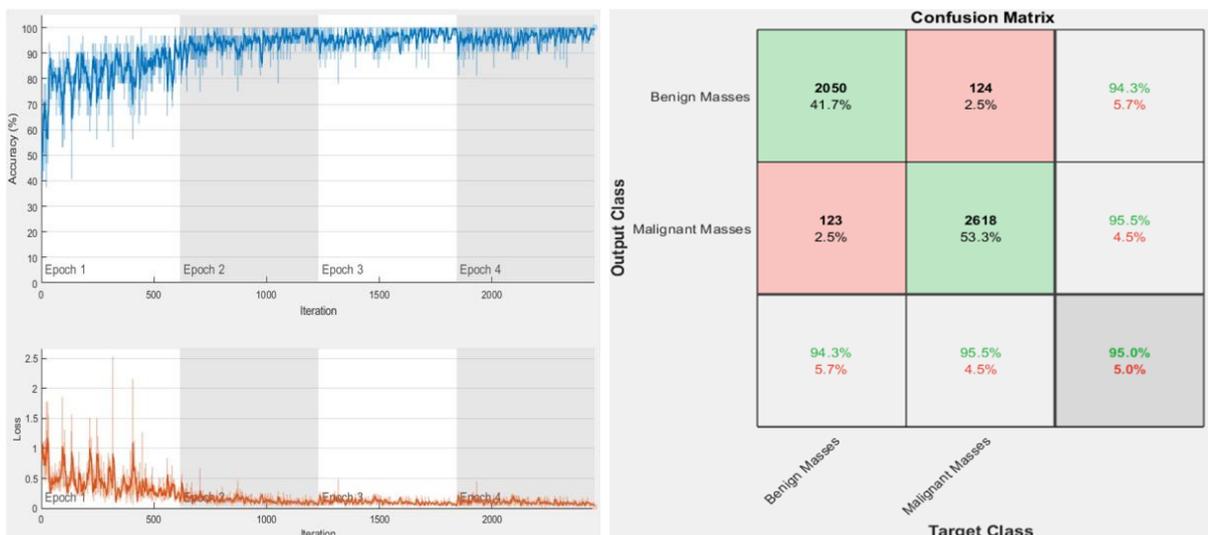


Figure 6. Best result for breast mass classification (ResNet18) with an adaptive learning rate.

Comparing Figure 5(d) and Figure 6 reveals that an adaptive learning rate offers enhanced stability, whereas a fixed learning rate can lead to issues like divergence or learning stagnation.

The adjustment of the learning rate based on model performance helps prevent these problems and ensures a stable learning process. Adaptive learning rates increase when the model is distant from convergence, facilitating exploration of the search space. They subsequently decrease the learning rate as convergence nears, allowing for more precise exploration around the optimal solution. Another noteworthy advantage is that adjusting the learning rate can prevent the model from overfitting the training data, thereby improving its capacity to generalize to new data. Additionally, this approach reduces the need for manual fine-tuning of the learning rate, which can be both labor-intensive and intricate.

Acknowledging the phenomenon where gradients diminish as convolutional neural networks deepen, ResNet addresses this challenge through shortcut connections, allowing the network to achieve greater depths. Moreover, ResNet networks provide the advantage of deep scaling, enabling the adjustment of architecture to the specific classification task at hand. This adaptability results in outstanding classification performance across various levels of image detail, including finer details, underscoring effectiveness in the analysis of medical images.

Table 10 provides a comparison of recent studies focusing on mammogram analysis. Our model stands out with a superior classification rate. In [30], the utilization of machine-learning algorithms raises concerns about potential divergence when dealing with a substantial volume of images. The authors of [32] employ a restricted sub-set of the Inbreast dataset (200 images), potentially resulting in a model tailored to the training data, but lacking robust generalization to new data. While surpassing the

"Using ResNet18 in a Deep-learning Framework and Assessing the Effects of Adaptive Learning Rates in the Identification of Malignant Breast Masses in Mammograms," S. Benbakreti, S. Benbakreti, K Benyahia and M Benouis.

achievements of [34] and [36], it is important to note that ResNet101 may pose challenges due to its resource-intensive nature. Moreover, the majority of contributions in the literature, as stated in the state of the art, use a single dataset. Our goal in summing the three datasets was to make the task more difficult. As shown in Table 10, the suggested model performs well, especially when the comparison is consistent [33], as a result of the authors' combination of the three datasets (MIAS+Inbreast and DDSM).

Table 10. Comparison between our model and related works.

Reference	Year	Used Dataset	Classification model	Accuracy
Zhang et al. [30]	2020	DDSM	SVM, Naïve Bayes, KNN	90.91%
Ting et al. [31]	2019	MIAS	CNNI+BCC	90.5%
Chougrad et al. [32]	2018	Inbreast (200 images)	VGG16	95%
Khartiga et al [33]	2022	Inbreast+DDSM+MIAS	CNN	92.27%
Wang et al [34]	2023	MIAS	SNSVM	94.1%
Fatima et al [36]	2023	CBIS+DDSM	ResNet101	94.5%
Proposed model	2023	Inbreast+DDSM+MIAS	ResNet18	95%

6. CONCLUSIONS

The goal of this research was to construct a convolutional neural network (CNN) model to distinguish between benign and malignant breast-cancer tumors. The used pre-trained models are AlexNet, InceptionV3 and ResNet18. After training on 80% of the dataset and testing on 20%, ResNet18 produced best classification result with an accuracy of 95%. The effectiveness of this model can be attributed to the incorporation of residual connections in ResNet18, which effectively addresses the issue of vanishing gradients during the training of deep networks. Furthermore, ResNet18 strikes a well-balanced equilibrium between classification performance and model complexity. It maintains a leaner profile compared to deeper architectures like ResNet50 while still delivering commendable performance. We also note that InceptionV3 achieved very good results (93.7%). This can be attributed to InceptionV3's architecture, which incorporates techniques, such as regularization and normalization, to reduce overfitting. Consequently, it minimizes the risk of over-adapting to the training data and improves generalization to new data. In addition, InceptionV3 employs factorized convolution, which reduces the number of parameters while maintaining high performance. This enhances its efficiency in terms of memory usage and computational resources. In summary, InceptionV3 is highly regarded for its ability to combine outstanding performance with a relatively lightweight architecture, making it a versatile choice for a variety of image-processing applications. We have also shown that adaptive learning rates enhance the efficiency, stability and performance of deep-learning models, making them a valuable asset in the training of neural networks, particularly CNN-based models.

In addition to investigating the impact of batch size and adaptive learning rate on the outcomes, this study provides insights into several key aspects. Firstly, it challenges the common notion that preprocessing and data augmentation are indispensable for achieving satisfactory results, suggesting that, under certain circumstances, these steps may not be imperative. Secondly, the integration of three reference datasets not only introduces a novel challenge, elevating the complexity of the classifier, but also opens avenues for exploring innovative approaches to enhance results. This approach aims to pave the way for more effective models that can truly serve as robust decision-support tools in real-world scenarios. Diagnostic-aid applications, especially those focused on identifying benign masses, prove to be invaluable tools in the medical domain. Their effectiveness is evident in their swiftness, precision and capacity to handle extensive datasets, such as mammograms. This enhances medical decision-making, facilitates early disease diagnosis and supports more targeted and efficient interventions. Additionally, these applications assist healthcare professionals in optimizing their time, especially during peak pandemic periods and help mitigate the risk of human errors. Nevertheless, it is essential to emphasize that the use of such applications should complement human expertise and not replace the clinical judgment of healthcare professionals.

REFERENCES

- [1] F. Bray, J. Ferlay, I. Soerjomataram, R.L. Siegel, L.A. Torre and A. Jemal, "Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries,"

- CA: A Cancer Journal for Clinicians, vol. 68, no. 6, pp. 394–424, DOI: 10.3322/caac.21492, 2018.
- [2] Y. Tan et al., "Tumor-derived Exosomal Components: The Multifaceted Roles and Mechanisms in Breast Cancer Metastasis," *Cell Death and Disease*, vol. 12, no. 6, DOI: 10.1038/s41419-021-03825-2, 2021.
- [3] M. Desai and M. Shah, "An Anatomization on Breast Cancer Detection and Diagnosis Employing Multi-layer Perceptron Neural Network (MLP) and Convolutional Neural Network (CNN)," *Clinical eHealth*, vol. 4, pp. 1–11, DOI:10.1016/j.ceh.2020.11.002, 2021.
- [4] J. Zuluaga-Gomez et al., "A CNN-based Methodology for Breast Cancer Diagnosis Using Thermal Images," *Computer Methods in Biomechanics and Biomedical Engineering*, "Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization", vol. 9, no. 2, pp. 131–145, 2021.
- [5] C. Gonçalves, J. Souza and H. Fernandes, "CNN Architecture Optimization Using Bio-inspired Algorithms for Breast Cancer Detection in Infrared Images," *Computers in Biology and Medicine*, vol. 142, DOI: 10.1016/j.combiomed.2021.105205, 2022.
- [6] D. A. Zebari et al., "Systematic Review of Computing Approaches for Breast Cancer Detection Based Computer Aided Diagnosis Using Mammogram Images," *Applied Artificial Intelligence*, vol. 35, no. 15, pp. 2157–2203, DOI: 10.1080/08839514.2021.2001177, 2021.
- [7] World Health Organisation, "Breast Cancer: Early Diagnosis and Screening," [Online], Available: <http://www.who.int/cancer/prevention/diagnosis-screening/breast-cancer/en/>, 2018.
- [8] W. Al-Dhabyani, M. Gomaa, H. Khaled and A. Fahmy, "Dataset of Breast Ultrasound Images," *Data Brief*, vol. 28, pp. 104863, DOI: 10.1016/j.dib.2019.104863, 2020.
- [9] D. Sheth and M. L. Giger, "Artificial Intelligence in the Interpretation of Breast Cancer on MRI," *Journal of Magnetic Resonance Imaging*, vol. 51, no. 5, pp. 1310–1324, DOI: 10.1002/jmri.26878, 2020.
- [10] T. K. Y. Tay and P. H. Tan, "Liquid Biopsy in Breast Cancer: A Focused Review," *The Archives of Pathology & Laboratory Medicine*, vol. 145, no. 6, pp. 678–686, 2021.
- [11] J. Kotsopoulos et al., "Tamoxifen and the Risk of Breast Cancer in Women with a BRCA1 or BRCA2 Mutation," *Breast Cancer Research and Treatment*, vol. 201, no. 2, pp. 257–264, DOI: 10.1007/s10549-023-06991-3, 2023.
- [12] Z. M. Colbert and P. Ramachandran, "Auto-segmentation of Thoracic Organs in CT Scans of Breast Cancer Patients Using a 3D U-net Cascaded into 2D PatchGANs", *Biomedical Physics & Engineering Express*, vol. 9, no. 5, DOI:10.1088/2057-1976/ace631, 2023.
- [13] G. Hamed et al., "Deep Learning in Breast Cancer Detection and Classification," *Proc. of the Int. Conf. on Artificial Intelligence and Computer Vision (AICV2022)*, pp. 322–333, DOI: 10.1007/978-3-030-44289-7_30, Cairo, Egypt, 2020.
- [14] N. S. Ismail and C. Sovuthy, "Breast Cancer Detection Based on Deep Learning Technique," *Proc. of 2019 Int. UNIMAS STEM 12th Eng. Conf. (EnCon)*, pp. 89–92, DOI: 10.1109/EnCon.2019.8861256, Kuching, Malaysia, 2019.
- [15] J. Wu and C. Hicks, "Breast Cancer Type Classification Using Machine Learning," *Journal of Personalized Medicine*, vol. 11, no. 2:61, DOI: 10.3390/jpm11020061, 2021.
- [16] S. Khan, N. Islam, Z. Jan, I. U. Din and Joel J. P. C. Rodrigues, "A Novel Deep Learning Based Framework for the Detection and Classification of Breast Cancer Using Transfer Learning," *Pattern Recognition Letters*, vol. 125, pp. 1–6, DOI: 10.1016/j.patrec.2019.03.022, 2019.
- [17] L. Shen, L. R. Margolies, J. H. Rothstein, E. Fluder, R. McBride and W. Sieh, "Deep Learning to Improve Breast Cancer Detection on Screening Mammography," *Scientific Reports*, vol.9, no. 1, p. 12495, DOI: 10.1038/s41598-019-48995-4, 2019.
- [18] Z. Han, B. Wei and Y. Zheng, "Breast Cancer Multi-classification from Histopathological Images with Structured Deep Learning Model," *Scientific Reports*, vol. 7, no. 1, p. 4172, DOI: 10.1038/s41598-017-04075-z, 2017.
- [19] D. H. Hubel and T. N. Wiesel, "Receptive Fields and Functional Architecture of Monkey Striate Cortex," *Journal of Physiology*, vol. 195, no. 1, pp. 215–243, DOI: 10.1113/jphysiol.1968.sp008455, 1968.
- [20] S. Li, L. Wang, J. Li and Y. Yao, "Image Classification Algorithm based on Improved AlexNet," *Journal of Physics: Conference Series*, vol. 1813, no. 1, p. 012051, 2021.
- [21] C. Szegedy et al., "Going Deeper with Convolutions," *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9, DOI: 10.1109/CVPR.2015.7298594, Boston, MA, USA, 2015.
- [22] A. Krizhevsky, I. Sutskever and G. E. Hinton, "Imagenet Classification with Deep Convolutional Neural Networks," *Proc. of the Advances in Neural Information Processing Systems (NIPS 2012)*, vol.25, DOI: 10.1145/3065386, Lake Tahoe, Nevada, United States, 2012.
- [23] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, DOI: 10.1109/CVPR.2016.90, Las Vegas, NV, USA, 2016.
- [24] Kaggle, "Breast Mammography Images with Masses," [Online], Available: <https://www.kaggle.com/datasets/tommyngx/breastcancermasses>, 04/10/2023.
- [25] F. Zhuang et al., "A Comprehensive Survey on Transfer Learning," *Proc. of the IEEE*, vol. 109, no. 1, pp. 43–76, DOI: 10.1109/JPROC.2020.3004555, 2021.

"Using ResNet18 in a Deep-learning Framework and Assessing the Effects of Adaptive Learning Rates in the Identification of Malignant Breast Masses in Mammograms," S. Benbakreti, S. Benbakreti, K Benyahia and M Benouis.

- [26] S. Benbakreti, M. Benouis, A. Roumane and S. Benbakreti, "Impact of the Data Augmentation on the Detection of Brain Tumor from MRI Images Based on CNN and Pre-trained Models," *Multimedia Tools and Applications*, DOI: 10.1007/s11042-023-17092-0, 2023.
- [27] E. M. Dogo et al., "A Comparative Analysis of Gradient Descent-based Optimization Algorithms on Convolutional Neural Networks," *Proc. of the Int. Conf. on Computational Techniques, Electronics and Mechanical Systems*, pp. 92–99, DOI: 10.1109/CTEMS.2018.8769211, Belagavi, India, 2018.
- [28] C. F. G. D. Santos and J. P. Papa, "Avoiding Overfitting: A Survey on Regularization Methods for Convolutional Neural Networks," *ACM Computing Surveys*, vol. 54, no. (10s), pp. 1-25, DOI: 10.1145/3510413, 2022.
- [29] K. Shankar, Y. Zhang, Y. Liu, L. Wu and C.H. Chen, "Hyper-parameter Tuning Deep Learning for Diabetic Retinopathy Fundus Image Classification," *IEEE Access*, vol. 8, pp. 118164–118173, DOI: 10.1109/ACCESS.2020.3005152, 2020.
- [30] X. Zhang et al., "Classification of Mammographic Masses by Deep Learning," *Proc. of the 2017 56th Annual Conf. of the Society of Instrument and Control Engineers of Japan (SICE)*, pp. 793–796, DOI: 10.23919/SICE.2017.8105545, Kanazawa, Japan, 2017.
- [31] F. F. Ting, Y. J. Tan and K. S. Sim, "Convolutional Neural Network Improvement for Breast Cancer Classification," *Expert Systems with Applications*, vol. 120, pp. 103–115, 2019.
- [32] H. Chougrad, H. Zouaki and O. Alheyane, "Deep Convolutional Neural Networks for Breast Cancer Screening," *Computer Methods and Programs in Biomedicine*, vol. 157, pp. 19-30, 2018.
- [33] R. Karthiga, K. Narasimhan and R. Amirtharajan, "Diagnosis of Breast Cancer for Modern Mammography Using Artificial Intelligence," *Mathematics and Computers in Simulation*, vol. 202, pp. 316-330, DOI: /10.1016/j.matcom.2022.05.038, 2022.
- [34] J. Wang, M.A. Khan, S. Wang and Y. Zhang, "SNSVM: SqueezeNet-Guided SVM for Breast Cancer Diagnosis," *Computers, Materials & Continua*, vol. 76, no. 2, DOI: 10.32604/cmc.2023.041191, 2023.
- [35] S. U. Rehman et al., "BRMI-Net: Deep Learning Features and Flower Pollination-controlled Regula Falsi-based Feature Selection Framework for Breast Cancer Recognition in Mammography Images," *Diagnostics*, vol. 13, no. 9, p.1618, DOI: 10.3390/diagnostics13091618, 2023.
- [36] M. Fatima, M. A. Khan, S. Shaheen, N. A. Almujaally and S. H. Wang, "B²C³NetF²: Breast Cancer Classification Using an End-to-end Deep Learning Feature Fusion and Satin Bowerbird Optimization Controlled Newton Raphson Feature Selection," *CAAI Transactions on Intelligence Technology*, vol.8, no.4, pp. 1374-1390, DOI: 10.1049/cit2.12219, 2023.
- [37] D. Shigemizu et al., "Classification and Deep Learning-based Prediction of Alzheimer Disease Sub-types by Using Genomic Data," *Translational Psychiatry*, vol.13, no. 1, p. 232, DOI: 10.1038/s41398-023-02531-1, 2023.
- [38] X. Liu, H. Wang, Z. Li and L. Qin, "Deep Learning in ECG Diagnosis: A Review," *Knowledge-based Systems*, vol. 227, p.107187, DOI:10.1016/j.knosys.2021.107187, 2021.
- [39] S. Al-Fahdawi et al., "Fundus-DeepNet: Multi-label Deep Learning Classification System for Enhanced Detection of Multiple Ocular Diseases through Data Fusion of Fundus Images," *Information Fusion*, vol. 102, p.102059, DOI:10.1016/j.inffus.2023.102059, 2024.
- [40] P. Roy, M. M. O. Chisty and H. A. Fattah, "Alzheimer's Disease Diagnosis from MRI Images Using ResNet-152 Neural Network Architecture," *Proc. of the 5th Int. Conf. on Electrical Information and Communication Technology (EICT)*, pp. 1-6, DOI:10.1109/EICT54103.2021.9733507, 2021.
- [41] R. Patnaik, P. S. Rath, S. Padhy and S. Dash, "Histopathology Colorectal Cancer Image Classification by Using Inception V4 CNN Model," *Proc. of the Int. Conf. on Robotics, Control, Automation and Artificial Intelligence*, pp. 1003-1014, DOI:10.1007/978-981-99-4634-1-79, 2022.
- [42] M. Dehghan Rouzi et al., "Breast Cancer Detection with an Ensemble of Deep Learning Networks Using a Consensus-adaptive Weighting Method," *Journal of Imaging*, vol.9, no. 11, p. 247, DOI: 10.3390/jimaging9110247, 2023.
- [43] A. B. Bagheri et al., "Potential Applications of Artificial Intelligence (AI) and Machine Learning (ML) on Diagnosis, Treatment, Outcome Prediction to Address Healthcare Disparities of Chronic Limb-threatening Ischemia (CLTI)," *Seminars in Vascular Surgery*, vol. 36, no. 3, pp. 454-459, 2023.

ملخص البحث:

سرطان الثدي مرض واسع الانتشار عالمياً، وهو مرض يصيب النساء في الغالب، لكنه قد يصيب الرجال كذلك. ويعدّ الكشف المبكر عن الإصابة بالمرض حاسماً في نجاعة العلاج؛ علماً بأنّ التصوير الشعاعي هو من بين الطرق المستخدمة في التشخيص. وتختلف التوصيات بشأن الصور الشعاعية المستخدمة للكشف عن الإصابة بسرطان الثدي حسب العمر، كما تختلف من بلدٍ لآخر.

تهدف هذه الورقة إلى استخدام نماذج تقوم على التعلّم العميق والشبكات العصبية الالتفافية بغرض تشخيص الإصابة بسرطان الثدي بناءً على صور الكتل الملتقطة للثدي. والجدير بالذكر أنّ تلك الكتل قد تكون إما خبيثة أو حميدة، مما يتطلب استراتيجيات تدخّل لكل سيناريو.

وتبيّن النماذج المقترحة في هذه الدراسة أهمية اختيار كلّ من حجم الحزمة ومعدّل التعلّم التكيّفي في التأثير على النتائج وتحسين دقّة التصنيف. وقد جرى استخدام مجموعة بياناتٍ مُدمجة من ثلاث مجموعات بيانات، مع حساب الدقّة عند استخدام كلّ منها، إلى جانب الدقّة عند استخدام مجموعة البيانات المجمّعة. وكانت أفضل دقّة تم الحصول عليها (95%) لمجموعة البيانات (RESNET18)، الأمر الذي يؤكّد ملائمة النظام المقترح لتشخيص الإصابة بسرطان الثدي.

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:



فريق دعم هيئة التحرير

ادخال البيانات وسكترير هيئة التحرير

المحرر اللغوي

إياد الكوز

حيدر المومني

جميع الأوراق البحثية في هذا العدد متاحة 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 الأميرة سمية للتكنولوجيا، شارع خليل الساكت، الجببية، عمان، الأردن.

صندوق بريد: 1438 عمان 11941 الأردن

هاتف: +962-6-5359949

فاكس: +962-6-7295534



جامعة
الأميرة سميرة
للتكنولوجيا
Princess Sumaya
University
for Technology



صندوق دعم البحث العلمي والابتكار
Scientific Research and Innovation Support Fund

المجلة الأردنية للحاسوب وتكنولوجيا المعلومات

ISSN 2415 - 1076 (Online)
ISSN 2413 - 9351 (Print)

العدد ١

المجلد ١٠

آذار ٢٠٢٤

JJCIIT

الصفحات	عنوان البحث
١٦ - ١	طريقة متعددة المجالات قائمة على تعلم الآلة لتمييز الكينونات المسماة لمحرّكات أنظمة التّحاور المتعلقة بالحجز باستخدام محوّلات ثنائية الاتجاه المدبّبة مسبقاً بشرى الصدر، رحمة الصدر، غيث عبندة، وإياد جعفر
٣١ - ١٧	تخطيط شبكات التّحديد في مجال التّردّدات الرّاديوية باستخدام خوارزمية العنقدة الهرمية التصاعديّة لامري سيّاد
٤٥ - ٣٢	خوارزميات للتوزيع والتخزين الآمنين لبيانات الفيديو كاندي أرشانا، و ف. كمكشي براساد
٥٧ - ٤٦	تصحيح أخطاء التّهجئة «الناعمة» باللغة العربية باستخدام نموذج تي ٥ محمد القراغولي، و غلا عارف جعفر
٧٣ - ٥٨	هوائي مبني على هندسة «كوخ» ثماني الأضلاع ذو تصميم مُدمج وبنية أساس مشوهة (DGS) للاستخدامات اللاسلكية في نطاق التّردّدات فائقة العرض (UWB) تيجاسويتا كوماري، أبو نصار غزالي، و أنوباما سيناباتي
٩٢ - ٧٤	نظام هجين لمعالجة العيوب في الصّور الملتقطة في أحوال الطّقس السيّئة ميلند فيجي بارسي، و دانيا برامود
١٠٧ - ٩٣	نموذج قائم على تكنولوجيا التّعلم العميق لتقييم أثر معدّلات التّعلم التّكيّفية في تحديد كتل الشّدي الخبثية من الصّور المستخدمة لتشخيص سرطان الشّدي سمية بن بكرتي، سمير بن بكرتي، كدا بن يحيى، و محمد بن ويس

www.jjcit.org

jjcit@psut.edu.jo

مجلة علمية عالمية متخصصة تصدر
بدعم من صندوق دعم البحث العلمي والابتكار