Jordanian Journal of Computers and Information Technology

JJCIT

Princess Sumaya University for Technology — جامعـة الأميــرة سميّــة للتكنولوجيا

Scientific Research Support Fund — صنـدوق دعم البحث العلمي

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

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

JJCIT

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	EDITORIAL BOARD SECRETARY
Haydar Al-Momani	Eyad Al-Kouz

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

JJCIT

EDITORS' NOTE

Dear authors, reviewers and readers of the Jordanian Journal of Computers and Information Technology (JJCIT).

It gives me great pleasure to welcome you to the Volume 8, Number 2 (June 2022) edition of the JJCIT for which I have acted as Editor-in-Chief (EIC). Firstly, thanks are due to my predecessor Professor Ahmad Hiasat, currently JJCIT senior editor, who successfully brought the JJCIT to be Scopus-indexed and more is to come as the JJCIT application is under review with Clarivate.

With the JJCIT indexing on Scopus, this allows the level of the Journal's metadata access to be increasingly higher by the user, because Scopus has features including citation, networking, research and score, so that it can help increase the Journal's citation index and reinforce Princess Sumaya University for Technology (PSUT) reputation globally.

I am very aware of the responsibilities that the EIC's role entails; therefore, I approach my new role with excitement and enthusiasm and will continue to observe and follow-up the quick and frequently dramatic changes in our field's research and technological landscape.

The JJCIT must adapt to this dynamic and rapidly changing publication scene. As EIC, I will work with the editorial and advisory boards to ensure that the JJCIT adapts appropriately and evolves to take full advantage of the fast changing world of publication and information dissemination. With you all, the JJCIT is accelerating its continued growth and peer review remains a vital component of our assessment of submitted articles.

I would like to thank the editorial and advisory boards again for their hard work and devotion, all of the authors for submitting their articles to the JJCIT and all of the reviewers for their comprehensive reviews of these articles. I am fortunate to be supported by a highly effective team, including Engineer Eyad Al-Kouz and Engineer Haydar Al-Momani in the editorial office who 'pulls the whole show together'.

I hope you will continue to provide the JJCIT your finest work.

Wejdan Abu Elhaija, PhD, DSc

Editor-in-Chief

CONSENSUAL BASED CLASSIFICATION AS EMERGENT DECISIONS IN A COMPLEX SYSTEM

Rabah Mazouzi[1], Malanga Kennedy[2], Cyril De Runz[3] and Herman Akdag[4]

ABSTRACT

In massive multi-agent systems that are used to model some complex systems, emergence is a key feature that allows to model high-level states of such systems. According to this perspective, the work we introduce in this paper entails the handling of emergence in massive multi-classifiers that we consider as complex systems. We aim to build a collaborative system for supervised data classification that we expect to provide better performance, compared to conventional classifiers. Modeled as a multi-agent system, the massive multi-classifier is composed of a high number of agents that are interconnected according to a given neighborhood. Each agent plays the role of a weak classifier. At the micro-level, the elementary interaction between agents consists of combining their respective classification results. Every agent, according to the majority vote rule, combines its result with those of its neighbors by taking into account their respective performances. This process is iterated continuously in a cyclic manner within the neighborhood of each agent. Therefore, a complex dynamic will be created within the system. After a certain time, this complex dynamic stabilizes, allowing the exhibition of an emergent structure that will be observed at the macro-level and is considered as a consensual class prediction for the data we want to classify. Obtained experimental results and the comparison with conventional classifiers show the potential of the approach to enhance classification and to be an alternative for classifier combination and aggregation.

1. INTRODUCTION

In automatic data classification, precision is a crucial aspect. Enhancing precision still remains an open issue. Since the First classifiers were proposed —which were simple and acted individually—, their accuracy has continued to increase asymptotically. We can see that there are many research works proposing new approaches to achieve more powerful classifiers [1]-[3]. Moreover, since the volume and dimensionality of data to be processed are increasing, it has become necessary to design and use more complex and sophisticated classifiers to process the data [4]. The collaboration of classifiers was among the former approaches, but also the most used to overcome the complex data classification problem. Such approaches propose collaborative architectures by combining multiple classifiers using various schemes [5]. Parallel and series combinations are both trivial schemes that make collaborate classifiers. According to the state-of-the-art in the field of classifier combination, it has been demonstrated that the combination of classifiers according to different schemas can produce better results compared to those obtained by the classifiers considered separately [6]-[9].

Since automatic classification of data is often involved in data processing systems, looking for and reaching acceptable accuracy in such systems remain an important need. So, it remains always necessary to propose new architectures of classifier combinations. Indeed, the complexity of processed data and their huge volumes often hide complex patterns and relationships, where classical Multi-Classifier Systems (MCSs) based on simple combinations cannot be considered to elucidate such relationships and patterns. So, we believe that the use of diversity provided on one hand by the multiplicity of classification algorithms and on the other hand on the diversity within the sets of training data, allows us to propose new unconventional and more sophisticated MCSs to deal with complex data [10]-[11]. Furthermore, in order to take advantage of the large volumes of training data, it is not appropriate to use multi-classifiers with a low number of elementary classifiers. Indeed, if the volume of training data assigned to each classifier is big, then it results in some conventional classification flaws, such as over-fitting and lack of generalization [12]. Therefore, it is necessary to conceive massive multi-classifiers

1. R. Mazouzi is with the Department of Research and Innovation, Umanis SA, Levallois-Perret, France. Email: rmazouzi@umanis.com, ORCID: 0000-0002-5945-5152.
2. M. Kennedy is with Dedan Kimathi Uni. of Techno., Kenya. Email: malangalanga@dkut.ac.ke, ORCID: 0000-0001-8435-5084.
3. C. de Runz is with BDTLN Team of LIFAT Lab., Tours, France. Email: cyril.derunz@univ-tours.fr
4. H. Akdag is with Paragraphe Lab., University of Paris 8, France. Email: herman.akdag@univ-paris8.fr

113

"Consensual Based Classification As Emergent Decisions in a Complex System", R. Mazouzi et al.

that can handle large masses of training data and have good performance in terms of generalization and accuracy. Moreover, the volume, variety and velocity properties with which big data are often defined represent challenges in the field of data mining [13]. According to our point of view, these factors increase the complexity of the system and diversity of classifiers. These are actually the two key aspects in the proposed MCS.

However, with a massive multi-classifier that consists of hundreds or even thousands of elementary classifiers, we cannot adopt conventional combination schemes, such as parallel or serial. It is therefore necessary to propose new collaborative architectures of classifiers that allow the use of a large number of elementary classifiers ensuring a coherent and parsimonious integration of the involved classifiers. Indeed, classical combinations of elementary classifiers through serial, parallel or hybrid combinations have not resulted into efficient classifiers, particularly when they are used in the context of big data. This is also true with classifier aggregation, such as boosting and bagging. In such a context, the classical techniques for classifier combination do not carry the high diversity of data. They also do not allow enough interaction between the elementary classifiers in order to produce a consensual classification. Furthermore, classical combination of classifiers and their aggregation cannot be used with massive multi-classifiers that involve hundreds or thousands of elementary classifiers. This is because it would be hard to ensure an explicit interaction within such large sets of interacting elements.

Emergence is an inherent property of complex systems. Despite its wide scientific use, the concept of emergence is not defined unequivocally, since it is used differently depending on the discipline. However, the concept of emergence can be defined as the appearance of a new property of the system at a higher level of observation, called macro-level; this phenomenon results from a dynamic interaction among entities at a micro-level of a complex system. Several computational paradigms are used to model complex systems [14]. Such systems are characterized by a large number of entities that locally interact producing a complex dynamic environment that in turn leads to emergent properties within the system. In this work, we propose an architecture of massive multi-classifier system that simulates a complex system. Each element in such a system represents a classifier located in a neighborhood of other classifiers and with which it interacts, exchanging data and decisions (the labels of the data point). Thus, a dynamic environment is created within the system which consists of perpetual exchange of information among classifiers located in neighborhoods. After a certain time, this leads to spreading of emerging decisions in the entire system.

Within a given neighborhood, classifiers proceed with a classical parallel combination with weighted majority voting [15]. Then, the decision taken by the agent at the center of its neighborhood is adopted by all the classifiers that are in the same neighborhood. Each agent will do the same work considering the neighborhood to which it belongs. Such collaborative approach can be considered as a scale-up of the classical combination and aggregation of classifiers. This allows involvement of a high number of elementary classifiers ensuring a dynamic way of making them interact resulting into a consensual classification.

The complex massive multi-classifier system is modeled as a Multi-Agents System (MAS) [16], where each agent has its own different classifier and interacts with neighboring agents. The interaction consists of exchanging decisions within the same neighborhood. A given agent combines the decisions of its neighbors with its own decision, then spreads in its neighborhood the result of the combination. In such a system —which is complex considering the large number of its interacting elements and the heterogeneity of the data—, we expect that a complex dynamic environment will result and lead —in the manner of self- organized systems— to the emergence of a structure within the system. This structure consists of clusters of neighboring classifiers' agents which have reached a consensus of classification (labeling). The ideal case is when a single cluster of agents having a global consensus emerges such that the cluster is composed of all the agents of the system. According to such a massive classification approach, the high volume and complexity of big data are well dealt with. The resulting classification method can be easily parallelized and implemented in cloud and high-performance computing systems.

The remainder of the paper is organized as follows: In Section 2, we provide a review of the literature concerning both collaborative systems of multi-classifiers and some emergence-based systems. This allows us to discuss various aspects of the emergence in complex systems. Section 3 is devoted to the proposed system in which we start by presenting its architecture and its components. Next, we show how the complex dynamic environment is created in the system and how structures emerge, representing

114

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 08, No. 02, June 2022.

the classification result observable in the system. In Section 4, we present the experimentation of our system by describing the experimental protocol, obtained results, analysis of the results and discussion. Finally, we conclude in Section 5 by summarizing our findings and highlighting some perspectives for this work.

2. RELATED WORK

As far as we know, few works have exploited emergence in the field of data mining. In fact, in such works, there are multiple challenges. For instance, it is a challenge to set local elements and patterns of interaction between elements so as to ensure the exhibition of an emerging phenomenon. On the other hand, it is also a challenge to detect and exploit an emerging phenomenon when it exists in a given outcome.

In the field of sociology, an agent-based modeling approach has been proposed by Y. Chen et al. in [17]. This modeling aims to study the concept of social capital; i.e., the benefits obtained by individuals through social interactions. These benefits can emerge in the form of social support, camaraderie, solidarity, influence …etc. This social capital should be measured in terms of the emerging structural properties generated by the links between homogeneous and like-minded individuals. In another work [18], the authors proposed a general framework based on social capital games for studying social structural pattern emergence.

A more general framework for the specification and simulation of emergence-based systems modeled as multi-agent systems was proposed by O. Paunovski et al. in [19] to control self-organized systems with emergence. They aimed at identifying in such systems some events specific to emergence and studying the causal relationships between the micro and the macro-levels in these systems. The proposed framework, which is software engineering-oriented, proceeds in two phases: In the first phase, the user proceeds by sequential and iterative refinement of the agent-based model, in order to ensure that the expected global behavior will be reached. This phase aims to detect some local elements that can influence the overall behavior of the system. In the second phase, a statistical correlation analysis allows to test if an emergence decision is observed within the system or not. Such observations allow the user to review some elements of the system in order to restart new modeling and simulation.

In a previous work, the same authors proposed a fuzzy approach for determining the herd forming in multi-agent system-based simulations [20]. The fuzzy reasoning is used to calculate mainly two values specific to individuals and groups within the system. The first expresses the membership of the individual to a given group and the second expresses the cohesion of this group.

In the tagging field, V. Robu et al. have studied the dynamic of tagging created in a collaborative system and how categorization patterns emerge from this activity [21]. In such systems, consensus on tags is reached and expressed as tag frequencies that follow a well-defined distribution law. From this, some structures emerge within the tagging system represented as graphs. The correlations within these graphs are used to extract tag vocabularies by partitioning them into sub-graphs formed by each of the correlated tags. The authors used the Kullback-Leibler distance for measuring the convergence of tag distributions. When this distance is close to zero, the convergence is assumed to have been reached. The time to be allocated to the system to reach a steady state, expressed by the number of tags produced for a given site, was also discussed. At the convergence of the tagging dynamic, the sub-graphs are constructed using a similarity criterion. Emerging vocabulary tags are then identified by using community detection algorithms.

To deal with problems related to data flow analysis in multimedia wireless sensor networks, Wang et al. proposed in [22] an agent-based model of a collaborative system for the classification of intruder targets, where audio information is collected by sensors and processed by statistical methods. Next, a step of classification of the characteristics of these data streams allows to provide the class of the observed target. To perform these treatments, multi-agent negotiation mechanisms preserving energy, which is an essential aspect in sensor networks, are specially designed to distribute the classification tasks among agents using the auction protocol. Individual decisions are combined in the manner of classical Multiple Classifier Systems (MCSs) in order to extend the life of the network and efficiently conduct the collaborative processing.

Among the possible schemes to implement a massive MCS, overcoming the problems associated with

115

"Consensual Based Classification As Emergent Decisions in a Complex System", R. Mazouzi et al.

the complexity of very large data, R. Mazouzi et al. proposed in [23] an architecture of a massive MCS modeled as a MAS and as an acquaintance network. In this system, the training data is distributed across a diversified set of classifier agents. Each agent is situated on a node of the network and is surrounded by a set of neighboring agents. The classification process begins with the arrival of the data to classify on a central node. The latter distributes the data across a set of agents designated according to their availability. Each agent labels the data to classify and combines the result with those of the agents in its neighborhood. In order to increase the accuracy of the system, the number of the classifiers involved in the combination can be increased by expanding the neighborhood (considering neighbors of neighbors).

Recently, Maystre et al. [24] have proposed a method to make several users interact in order to collaboratively classify a set of items assuming that the labels are corrupted by noise. Their method is based on a structured probabilistic model that relates the interaction user-item and the noisy labels to the items to classify. Such interaction allows collaboration within the set of users in order to infer the correct label of a given item. In order to classify data streams in the context of Internet of Things, Sun et al. [25] proposed a Misclassification-Aware Collaborative Classification Algorithm (MACCA) that makes two modules to collaborate: the misclassification judgment module and the decision one. Such a strategy is proposed as an alternative of the vote technique, which is considered as degrading results by the authors.

There are few systems that use MAS-based architectures for constructing multi-classifiers. The works proposed according to this paradigm aim at generating new schemes for combining classifiers, different from conventional schemes, based on the simple serial and/or parallel combination. However, no system to our knowledge has addressed the problem from the point of view of complex systems, where emergence can be exploited to differently represent the classification results.

3. A COMPLEX SYSTEM FOR DATA CLASSIFICATION

3.1 Overview

We propose in this paper an emergence-based approach for supervised classification. The approach involves using a complex system paradigm modeled as a multi-agent system with the aim of building a massive multi-classifier for classifying large and complex data. So, the proposed system can be used in the context of Big Data. The multi-agent system consists of a large number of agents organized in a neighborhood system, where each agent is located within several neighbors' agents. The neighbors' agents number varies within a given range. An agent has knowledge about its neighbors and is able to interact with them. It also has a classifier, trained by a dedicated training data subset.

After the training phase of all the classifiers —in order to have the most possible diversified classifiers by using different training subsets—, the system is ready to receive the data item to be classified. The latter, when available, is delivered to all agents in the system. For centralized implementation of the proposed system, this is done immediately by initializing the data to classify among all agents. For a physical distributed implementation, a delivery mechanism must be considered. It may consist in placing the data on any agent, then allowing agents to deliver to their neighbors. Thus, the data is progressively propagated in the system and finishes by reaching all the agents.

An agent begins with classifying a data point using its own classifier. Then, cycle after cycle, it reclassifies the data point, considering the results of classification provided by its neighboring agents. The reclassification consists of a combination of local results, taking into account the accuracy and the performance of every classifier involved in this combination. For this, the classification method adopted shall allow to re-inject results of the previous classification as prior knowledge to the new classification. Considering a given neighborhood, the classification is performed in this neighborhood according to the classical parallel combination with weighted majority vote [26]. The result of the combination may be considered by the agent in question or not, according to the certainty of classification and the performance of neighboring classifiers.

According to this pattern of interaction, classification results in a given neighborhood can spread in the system, because a neighbor agent is in turn a neighbor to other agents in other neighborhoods. The propagation evolves according to the quality of classification. Indeed, if the data point is well classified, then there is a tendency that the neighboring agents spread the result beyond the neighborhood.

However, if the classification quality is lower, there is a tendency that the agents beyond the neighborhood reject this result. The fact of repeating such interactions leads to creation of a complex dynamic environment and after a given time, the dynamic stabilizes on emerging patterns, representing the overall result of classification within the whole system. Ideally, a cluster of the majority of the agents producing the same result is formed. In other cases, separate clusters can be formed and a decision mechanism must be expected on how the result of classification must be retrieved. The worst situation happens when a cyclical dynamic environment —on the edge of chaos— remains in the system. In this case, no overall result can be observed and considered.

3.2 The Classifier Agent

It represents the basic element of the system the task of which is to classify the data presented at the input and interact with its neighbors in order to lead the system to a global consensus in terms of classification. The interaction consists of exchanging the classification results by sharing its own results and receiving the results of its neighbors, then combining them according to a given method of combination. Figure 1 shows the architecture of the used agent. The agent has two roles:

3.2.1 Training

The agent proceeds to train its classifier by using a training sub-set extracted by random sampling from a global training dataset. Let DS be the overall training dataset. By using statistical sampling, DS is partitioned into N sub-sets DS_i, i = 1..N.

The sampling method is based on some probability distribution selected according to the size of the overall training dataset, the nature of data it contains and the application field. In our case, DS is homogeneous; so, the uniform probability law may be used for its stochastic partitioning (Algorithm 1).

Algorithm 1 sampler

> **procedure**   Sampler(DS, N)   ▷ dataset and agents' count
>> $k \leftarrow \frac{M}{N}$
>> **for** $i \leftarrow 1,N$ **do**   ▷ for each agent
>>> $DS_i \leftarrow \phi$
>>> **for** $i \leftarrow 1,N$ **do**   ▷ for each sample
>>>> $Pr \leftarrow Random$
>>>> $idx \leftarrow M \cdot Pr$
>>>> $DS_i \leftarrow DS_i \cup DS[idx]$
>>> **end for**
>> **end for**
> **end procedure**

We have considered "not disjunctive" sub-sets (sampling with replacement), so that the resulting sub-sets overlap with each other. Therefore, the different agents will be close in terms of classification results. Indeed, two agents which share some elements in their respective training subsets tend to produce similar results, thus promoting consensus while computing. This allows controlling the degree of diversity in the set of classifiers, which in the extreme diversity case hinders the convergence of the dynamic within the system.

3.2.2 Classification

This role is performed by the agent in two steps:

1) The calculation of the initial class, using its own classifier.
2) The re-calculation of the class using the classification results of neighbouring agents according to a combination rule. This calculation is iterated while the dynamic of the system has not been stabilized.

In order to be able to re-introduce the combinaion result in the subsequent calculation step, we opted for the Naive Bayesian classifier [27].

Let k classes $C_1, ..., C_k$ and X be the data point to classify; the class to consider, i.e., Label, corresponds

to the Maximum A posteriori Probability (MAP), according to the Bayes probability law:

$$P(C/X) = \frac{P(X/C).P(C)}{P(X)} \tag{1}$$

In our case, P (X) is constant for the whole data, thereby maximizing P (C/X) is equivalent to maximizing P (X/C)·P (C) and in this case, the resulting class Label may be expressed as follows:

$$Label = argmax_{j,1..k}(P(C_j/X).P(C_j)) \tag{2}$$

This represents the initial calculation made by any agent in order to calculate the first probability vector according to the Bayes probability law. The agent, in subsequent cycles, should always keep a probability vector, which corresponds to the best classification, since the start of the calculation.

Combination Rule

The combination within a given agent uses the classification results from its neighbors' classifiers. These classification results are considered by the agent in question as a probability vectors, where each one corresponds to a neighboring agent. Let $Pr_j^{i(X)}$ be the probability corresponding to the class C_j , calculated by the classifier i and let b_i be the classifier weighting factor i, calculated from the error rate obtained at the training step using a test dataset. This weighting factor expresses the degree of importance of the classifier, compared to neighboring ones.

$$g_j(X) = \sum_{i \in Neighbors(A)} b_i \cdot Pr_j^i(X) \tag{3}$$

By this expression, we calculate the elements of the probability vector corresponding to the combined weighted prediction according to the majority vote rule. After the calculation, the agent compares the quality of the new combination with that it keeps as the last better classification quality. The new classification combination is applied only if it is better than that stored at the agent.

When the agent adopts a new vector of probabilities, it will use it as a vector of classes prior probabilities ($P(C)$) in the next calculation cycles. In both cases, the agent starts a new calculation cycle after waiting a given time, required to insure that its neighbors have performed their own cycles.

The pseudo-code in Algorithm 2 represents the cycle (combining process) of an agent A:
Let *Pr(X)* be the set of the vectors of probabilities corresponding to the agents neighboring the agent *A* and *A.Prediction* be its own vector. The following pseudo-code calculates the new probability vector as a result of combination with the neighboring vectors:

Algorithm 2 Agent cycle

    procedure A.COMBINE($Pr(X)$)
        for $j \leftarrow 1, Prediction.size$ **do**
            $g_j(X) \leftarrow 0$
            for $i \leftarrow 1, A.Neighbors$ **do**
                $b_i \cdot Pr_j^i(X)$
            end for
        end for
        $PrMax \leftarrow argmax(g_j(X))$
        if $PrMax >$ argmax($A.Prediction$) **then**
            $A.Prediction = g(X)$
        end if
    end procedure

The class kept by the agent corresponds to the maximum of probabilities and its probability represents the quality of the classification of the data point X by the agent. Indeed, the latter updates its classification result according to the combination rule if and only if the new quality (MAP in the vector *PrMax*) is greater than the quality it keeps; namely, *max(A.Prediction)*.

Synchronization

When an agent is about to carry out a re-calculation of combination, it invokes its neighbors to get their

118

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 08, No. 02, June 2022.

classification results. It is possible that some of these agents may still have not achieved their current cycle since the last iteration. Hence, their results are the same as used by the agent in question during the previous cycle. But, even if that's the case, this does not adversely affect the calculation. For the next iteration, the calculation will be carried out with the new results of neighboring agents when they are ready.

3.3 The Multi-agent System

As shown in Figure 2, the complex system is built as a massive multi-agent system. First, each agent within the system is trained by a randomly selected sub-set of the whole training data. At this stage, agents are independent, since they do not interact with each other. After agents are created and trained, the neighboring relations are randomly created, where each agent is linked to a set of other agents. This ensemble forms the local neighborhood, where decisions are shared and adjusted. After the dynamic environment of the system is stabilized, the emergent decision (final classification) is provided as an output of the whole system. In the next sub-sections, we provide further details of how such stages are initialized and executed in the multi-agent system.

3.3.1 Creation

A set of N agents are created to form the multi-agent system. Each agent; let it be A, at its initialization,   selects a set of NV_A agents that form its neighborhood. This number is arbitrarily chosen by the agent in question belonging to the interval [Min_V, Max_V] (Algorithm 3). The number of neighboring agents   defines the density of acquaintances in the system, where the arbitrary choice of this number generates an asymmetrical system. The asymmetrical property allows the multi-classifier system to acquire the aspect   of heterogeneity necessary to explore vast spaces of states in search of the best consensual classification.

Algorithm 3 Create Neighborhood

procedure   A.CreateNeighberhood(Agents)   ▷ agents set
    $A.Neighbors \leftarrow \Phi$
    $NV_A \leftarrow MinV + random(MaxV - MinV)$
    for $i \leftarrow 1, NA_A$ **do**
        $idV \leftarrow 1 + random(Agents.size)$ ▷ agent id$\neq$A.id
        $A.Neighbors \leftarrow A.Neighbors \cup \{Agents(idV)\}$
    end for
end procedure
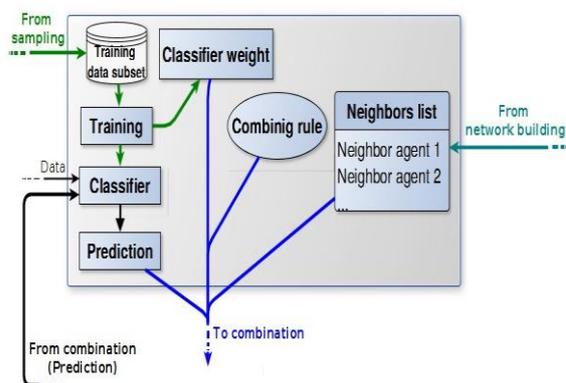
Figure 1. Agent structure.
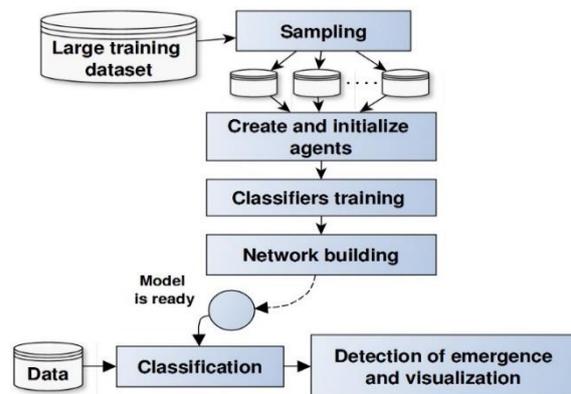
Figure 2. System building.

3.3.2 Execution

After the system is created, the data point to classify is submitted to all agents in the system. In the First step, each agent performs the initial classification of the data point using the Bayesian classifier, trained by its own training sub-set. Then, agents initiate a long interaction phase, where in each cycle, an agent carries out the re-classification of the data point by the combination of its result with those of its neighboring agents. This is repeated by all the agents of the system until a convergence of the system

119

"Consensual Based Classification As Emergent Decisions in a Complex System", R. Mazouzi et al.

dynamic occurs. This convergence is indicated by the fact that any agent maintains a steady state, expressed by the class that it holds and becomes constant over time.

Figure 3 shows an example of an agent surrounded by its neighbors. The central agent *A* retrieves the classification results of a given data point *X* from the neighboring agents, respectively *B*, *C*, *D*, *E* and *F*. The calculated combination is kept within the agent A. The neighboring agents *B* . . . *F*, when they perform their cycles, retrieve in their turn the classification result of the agent *A*.

3.4 Collaborative Classification

The classification according to the presented scheme can be considered as consensual classification. Each agent is under the influence of two trends:
- Its own classification (initial one), obtained from its local classifier and
- Its neighborhood, where the classification results are provided from other different classifiers. Therefore, over time, some dominance in the system —in term of classification results— tends to spread to the whole system.

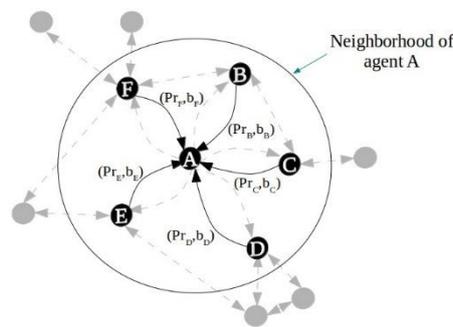Figure 3. Local interaction of an agent.

The complex dynamic environment that is created and maintained over time in the system tends to converge to a stable state. This state represents a consensus of classification within the population of agents. The ideal case is that all agents are stabilizing on the same result. An alternative case is to have clusters of agents that led each one to the same classification result. Each cluster is naturally contiguous due to the local interaction of agents within their respective neighbors. Otherwise, the system remains in an unstable state, on the edge of chaos, where no even partial consensual results of classification can be observed.

The classification thus made by consensus of classifiers, observed by the emergence of stable structures within the system, represents a compromise decision between the different agents. Indeed, it is certain that if a given data point that can be conveniently classified by some classifier does not necessarily get the same result with other classifiers, since they have different training sub-sets. The aim of our approach is to achieve the best compromise between the different agents of the system, so that agents which have contributed significantly through their classification quality dominate the final decision. In fact, there is a steady state in the space of states characterized by a low dynamic environment exhibiting the consensual classification. This is considered as the objective function to be optimized.

3.5 Convergence of the Complex Dynamic and Emergence of Consensus

The complex dynamic environment, in the sense of our approach, consists of the degree of variation of the states (calculated classes) within the population of agents over time. Initially, the dynamic should be high, since the state of each agent is the result of its own decision, considering its own knowledge (obtained from its local initial classification model). Then, some agents begin to change their states under the effect of the decision combination rule. If there is an optimum of the dynamic, expressed by a classification consensus, this dynamic tends to be reduced and stabilizes over time.

To detect the convergence of the dynamic interaction of agents, we observe two phenomena:
1) It is considered that there is convergence if all agents (or the majority of them) keep their states unchanged for several execution cycles. When the dynamic environment stabilizes, the system stops and classification result can be retrieved.

2) By using some global entropy quantification to measure the degree of variation of the overall system state. This global state is the superposition of all states of the agents within the system. In the literature, several studies have used entropy to measure activity in dynamic systems. Indeed, when the system becomes homogeneous, its entropy tends to be minimal [28]. The homogeneity corresponding to the absence of any structure within the system is called in this case isotropic. In digital systems, such as ours, we specifically use the Shannon entropy [29]. If the calculated entropy is low enough, we assume that there is a global consensus within the population of agents and the authenticity of the obtained results, depends on the value of the entropy.

In complex systems, the dynamic environment is something that is observed by an external observer. The convergence of the dynamic that can lead to the emergence of certain stable structures within the system can be detected only by this observer. In our case, we must be able to perform a number of steps to determine whether there is convergence or divergence in the sense of emergent stable structures within the population of agents.

To do this, we arbitrarily choose a small number of agents in the system and follow the evolution of their states with the aim of detecting convergence of the dynamic within this sub-set of agents. This idea is inspired by the Monte-Carlo method [30], which involves estimating properties of a system by measuring properties related to a sub-set of its elements drawn randomly. In our case, each agent in the considered subset is observed in order to see whether its state becomes invariant over time. This state corresponds to a class that the agent has by calculating the maximum *a posteriori* probability according to Bayes law. The system continues evolving until the states of the various agents in the observed sample become invariant. In this case, the dynamic environment is considered to have stabilized, then we proceed to the entropy measure to decide whether there is consensus within the population of agents or not.

3.5.1 Sample Selection and Test of Convergence

We consider 10 % of the population of agents as a sample for measuring the convergence of the dynamic. This number is assumed sufficient to estimate the overall state of the system. Let E be the set of agents representing the observed sample. To test the stability of the state of each agent A within the sample, we need to perform certain calculations. For each agent A, an attribute, called Invariant, is initialized to False, given that all agents are not invariant at the beginning of the dynamic.

Within an agent, at each execution cycle, the pseudo-code in Algorithm 4 is executed; so, after several cycles, the value of Invariant is set to True when the corresponding agent executes several cycles (CyclesThreshold).

Algorithm 4 Test of agent's state invariance

```
if A.NewState=A.CurrentState then
    Inc A.CyclesCount
else
    A.CyclesCount ← 0
end if
if A.CyclesCount>CyclesThreshold then
    A.Invariant← True
end if
A.NewState=A.CurrentState
```

CyclesThreshold, which represents the number of cycles to consider before deciding whether the dynamic stability is reached or not, is a system parameter defined experimentally using a testing dataset.

After a certain amount of time to allow the system to move towards a convergence state, tests are performed periodically by an observer entity of the system. This involves testing whether all the states' agents in the considered sample are invariant or not (all set to *True*). The observer pseudo- code to test the overall convergence is introduced in algorithm 5. If the time allowed for the system to stabilize; i.e., *MaxTime*, has elapsed without convergence, we can deduce that the system remains in permanent instability and a convergence state is not possible for the data point to be classified. The maximum period of calculation will also be set experimentally.

3.5.2 Entropy Calculation

Equation 4 expresses the entropy, corresponding to a system with M states, where P_i represents the probability that the system is in the state i. We use the value of entropy to identify the presence or absence of structures in the population of agents. Typically, the presence of clusters of homogeneous agents (leading to the same class), corresponds to a low entropy. In contrast, the divergence of all the agents in terms of classes corresponds to high values of entropy. Ideally, if all agents have concluded to the same class, we will have the lowest value of entropy.

Algorithm 5 System overall convergence

$Convergence \leftarrow False$
while $ComputingTime < MaxTime$ **and** $Convergence = False$ **do**
    if $\forall i, E(i).Invariant = Ture$ **then**
        $Convergence \leftarrow True$;
    end if
    wait($Delay$)
    $ComputingTime \leftarrow ComputingTime + Delay$
end while
if $Convergence = True$ **then**
    Computing the entropy
else
    Non stability $\Rightarrow$ Consensus failed
end if

$$H = -\sum_i P_i . Log P_i \tag{4}$$

Let A be a given agent, with NV_A neighbors. The entropy in the neighborhood of this agent is expressed as follows:

$$H_A = -\sum_{i=1}^{C} P_A^l . Log P_A^l \tag{5}$$

where C is the set of classes and P_A^l the probability that the agent A concludes to the class l. We estimate P_A^l by calculating the relative frequency of class l in the neighborhood of the agent A.

$$P_A^l = \frac{Number\ of\ agents\ concluded\ in\ class\ l}{NV_A + 1} \tag{6}$$

H_A is null if all agents have the same state, $\exists P_A^c = 1, \forall k, k \neq c, P_A^l = 0$; in this case:
$-\sum P_A^l log P_A^l = -P_A^c log P_A^c = 0$, Which corresponds to the minimal value of H_A.

The entropy at an agent is high if there is a divergence in the different classes concluded in its neighborhood. It is low in the opposite case and null if all the agents in the neighborhood conclude to the same class.

H_A is high if $P_A^l, l = 1...C$ are uniformly distributed.
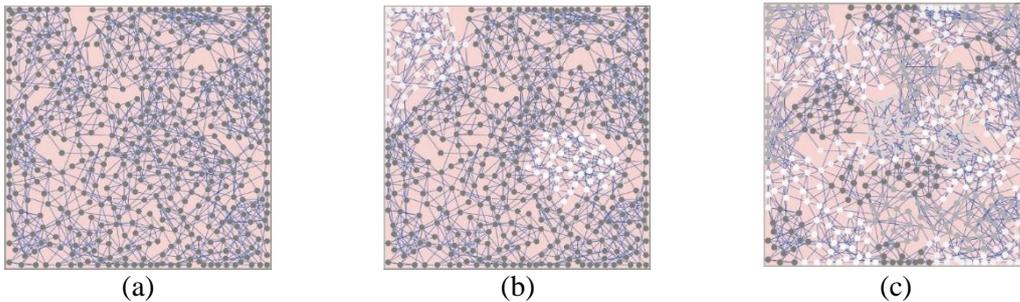

(a)   (b)   (c)

Figure 4. Clusters of agents according to H: (a) The ideal case where H is minimal: presence of one cluster (representing the emergent consensual decision); (b) H is low: presence of clusters, but we can retrieve a dominant class from the biggest cluster; (c) H is high: no emergent structure.

Global Entropy

Considering the entropy in each agent of the system, the overall entropy of the system is the sum of the entropies of the different agents: $H = \sum_A H_A$

The use of the thresholds for the global entropy H allows determining in which state the classification system has been stabilized. Figures 4 (a), (b) and (c) illustrate some examples of convergence with different system entropies. Various colors represent the difference in terms of classes concluded by each respective agent, corresponding to the data point to be classified.

4. EXPERIMENTATION

We have performed experimentation to demonstrate the significance of the proposed classification approach. In the following experimental analysis, we will show that results are obtained according to emergent consensus within the population of agents and that leveraging emergence in such complex system results in improvement of the classification results.

First, we consider various training datasets. From each training set, we will extract N non-disjoint training sub-sets and link each one to an agent within the multi-agent system. Such a way of selecting training data ensuring the desired diversity is a key aspect for Big Data classification. To do this, we consider several agents selected randomly within their population of agents. Each selected agent with its neighboring agents are considered as a classic parallel multi-classifier. We first calculate the accuracy of each classifier separately; then, we determine accuracy of results from the combination according to a weighted majority vote of each agent with its neighbors. Finally, we compare the obtained accuracy with that obtained by detection of emergent classes.

At the current state of our work, we have proceeded for experimenting our approach by an endogenous comparison, where results obtained according to the proposed method are compared to those obtained with conventional classifiers. We used the Naive Bayesian classifier, because it provides the quality of labeling (as the posterior probability) of a given data item. This is a key attribute necessary to make interacting agents and adjusting their decisions.

In a first experiment, we put 500 agents in interaction to classify a test set of 10,000 instances from the well-known KDDCup99 dataset on intrusion detection, which contains about 5 million instances, described by 43 attributes [31]. The comparison of: First, the five best average accuracy rates obtained by classifiers, taken separately; second, the five best sets of classifiers combined with their direct neighbors; and third, the accuracy rate achieved by the system of emergent classes, is as follows:

- 84.11% of accuracy rate on average for the 5 best classifiers.
- 84.88% of accuracy rate on average for local sets of classifiers (neighborhood).
- 86.85% achieved by the overall system (with emergence).

This result confirms our hypothesis that with a complex system with "relevant" interactions at the local level, the labels that emerge at the global level within this system are overall better than all the results of individual classifiers or local sets of classifiers. On the other hand, the weak point of the complex system-based method is mainly related to the computation time.

So, in terms of execution time, the classification time for the entire test sub-set is 7920 seconds, which is almost 4.5 times longer than classifying with a three-neighborhood level (considering neighbors of neighbors) and almost 8 times longer than using one level of neighborhood. The deployment of the experience to a real Cloud architecture could reduce the time of classification, but this does not seem fully compatible with big data labeling, because:

- The execution time is longer,
- The entire infrastructure is used to classify each data point.

Additionally, we should note that significantly better accuracy rates can be found in other work dealing with KDDCup99. This can be mainly the consequence of the choice of the type of classifier (Naive Bayes) which has relatively low individual performances on the considered dataset. Nevertheless, it is the most suitable algorithm for our approach, because it provides some parameters necessary for agent interaction. However, the goal of our experiments was to confirm the improvement brought by the approach based on a set of classifiers compared to classifiers taken separately and the improvement

brought by our approach based on the emergence within the global system, compared to that based on small sets of classifiers.

Table 1. Accuracy rates (Agents, Neighborhoods, Global) - Connect-4.

	Top three accuracy rates (%)	
	Agents	**Neighborhoods**
#1	55	75,4
#2	53,6	73,9
#3	52,1	72,5
Average	53,5	73,9
Global	**78,3**	

Table 2. Accuracy rates (Agents, Neighborhoods, Global) – Covertype.

	Top three accuracy rates (%)	
	Agents	**Neighborhoods**
#1	63,1	75,9
#2	62,8	75,3
#3	62,1	75,1
Average	62,8	75,4
Global	**76,7**	

In order to consolidate the results achieved with the KDDCup99 dataset, we have used less known other datasets: Connect−4 [32] and Covertype [33]. The following is a brief description of each one:

- *Connect−4* is a game, where each of the two players tries to line up four of their pieces in a grid of 6 rows and 7 columns. The strategy of the game is to try to line up four of one's own pieces, but at the same time prevent the opposing player from having a row of four of his/her pieces. Each space in the game grid represents an attribute of the dataset and each instance in the dataset represents the legal positions corresponding to the first 8 moves of a game for which no player has yet won. The class represents the result of the game *vis-à-vis* the first player (won, lost or drawn); this result will be obtained if both players continue to play "perfectly". The dataset contains 67557 instances (games), 42 attributes and 3 classes.

- *Covertype* dataset is composed of geospatial data created by the USFS Forest Inventory and Analysis (FIA) and the Remote Sensing Applications Center (RSAC) to show the extent, distribution and composition of forest cover in the United States. It contains 581,012 instances, 54 attributes and 7 classes. Each instance contains data for a 30 x 30 meter cell and the class represents the type of forest cover.

Tables 1 and 2 represent results recorded for the *Connect−4* and *Covertype* datasets, respectively. We have noted the three best accuracy rates considering the results obtained at the agent level and at the neighborhood level. The overall accuracy rate achieved by all the agents with a consensus on the classification is also shown.

The obtained results show the improvement in accuracy rates from the agent and its neighborhood to the overall system result (Figure 5). The improvement recorded by agent sets in the neighborhood is explained by the principle of the classical combination of classifiers (on average 73,9% for *Connect–4* and 75,4% for *Covertype*). We also note the improvement of accuracy rate at the level of the overall system (78,3% for *Connect-4* and 76,7% for *Covertype*), achieved by the combination of results within the global population of agents having reached consensus for labeling (emerging classification).

These results confirm our hypothesis about the potential of the interaction of agents (classifiers) in a massive system for improving data classification accuracy. However, certainly there is still some more work to be done in studying the dynamic environment within such systems so as to understand how the parameters inherent to these systems influence the quality of the results.
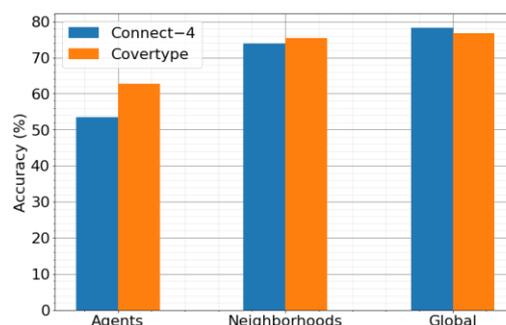

Figure 5. Accuracy rates improvement for *Connect-4* and *Covertype*.

124

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 08, No. 02, June 2022.

5. CONCLUSION

In this work, we proposed a consensual approach for data classification. The approach is intended to be adapted to complex data. Such data which can be of very high volume, of high dimensionality and possibly distributed can be handled efficiently using our approach with emphasis on both the quality of classification and the scalability of the solution. We used the paradigm of complex systems modeled as multi-agent systems to design a massive multi-classifier system. We have qualified it as massive, because it is assumed to contain a very large number of elementary classifiers put in interaction. The interaction generates phenomena relating to complex systems, such as the emergence of structure at the macro-level of the system. The emergent structure we desire is the consensual decision that the population of a majority of agents may have. The timeline is arrived at when agents that belong to local neighborhoods communicate their classification results. Then, over time, a complex dynamic environment sets within the system, which tends to converge on an emerging structure that represents a classification consensus within the population of agents. Finally, we retrieve this consensual decision as the label of the treated data point. Our approach is not suitable for a system that needs real-time processing. This is because in such a system, the time needed for the complex dynamic environment to stabilize so that the phenomenon of emergence appears may be too long. This is not desirable for a real-time processing system.

In the future perspective of this work, it is essential to push the experiments further in order to improve the proposed approach. This will enable better understanding of how to refine the parameters of the system and control the created dynamic environment in order to achieve better classification results.

REFERENCES

[1] S. Umadevi and K. S. J. Marseline, "A Survey on Data Mining Classification Algorithms," Proc. of the IEEE Int. Conf. on Signal Process. Comm. (ICSPC), pp. 264-268, Coimbatore, India, 2017.

[2] S. Minaee, N. Kalchbrenner, E. Cambria et al., "Deep Learning-based Text Classification: A Comprehensive Review," ACM Computing Surveys, vol. 54, pp. 1-40, 2021.

[3] A. H. Mohd Aman, E. Yadegaridehkordi, Z. S. Attarbashi, R. Hassan and Y. Park, "A Survey on Trend and Classification of Internet of Things Reviews," IEEE Access, vol. 8, pp. 111763- 111782, 2020.

[4] L. F. Cordeiro Robson, C. Faloutsos and C. Traina Jnior, Data Mining in Large Sets of Complex Data, ISBN-13: 978-1447148890, Springer Publishing Company, Incorporated, 2013.

[5] L. I. Kuncheva, "Combining Pattern Classifiers: Methods and Algorithms", Wiley-Interscience, 2004.

[6] X. Dong, Z. Yu, W. Cao, Y. Shi and Q. Ma, "A Survey on Ensemble Learning," Frontiers of Computer Science, vol. 14, no. 2, pp. 241-258, 2020.

[7] O. Sagi and L. Rokach, "Ensemble Learning: A Survey," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 8. No. 4, DOI: 10.1002/widm.1249, 2018.

[8] F. Yucalar, A. Ozcift, E. Borandağ and D. Kilinç, "Multiple-classifiers in Software Quality Engineering: Combining Predictors to Improve Software Fault Prediction Ability," Engineering Science and Technology: An International Journal, vol. 23, no. 4, pp. 938-950, 2020.

[9] S. Tulyakov, S. Jaeger, V. Govindaraju and D. Doermann, "Review of Classifier Combination Methods," Proc. of Machine Learning in Document Analysis and Recognition, Part of the Studies in Computational Intelligence Book Series, vol. 90, pp. 361-386, 2008.

[10] L. I. Kuncheva and C. J. Whitaker, "Measures of Diversity in Classifier Ensembles and Their Relationship with the Ensemble Accuracy," Machine Learning, vol. 51, no. 2, pp.181-207, 2003.

[11] D. Gacquer, On the Usability of Diversity When Training Multiple Classifier Systems. Application to the Detection of Hazardous Smokes on Industrial Facilities, Thesis at University of Valenciennes and the Hainaut-Cambresis, France, 2008.

[12] R. Roelofs et al., "A Meta-analysis of Overfitting in Machine Learning," NeurIPS Proceedings, [Online], Available: https://papers.nips.cc/paper/9117-a-meta-analysis-of-overfitting-in-machine-learning, 2019.

[13] D. Che, M. S. Safran and Z. Peng, "From Big Data to Big Data Mining: Challenges, Issues and Opportunities," Proc. of the 18th Int. Conf. DASFAA, vol. 7827, pp. 1-15, Springer, 2013.

[14] M. A. Niazi, "Introduction to the Modeling and Analysis of Complex Systems: A Review," Complex Adaptive Systems Modeling, vol. 4, Article no. 3, DOI: 10.1186/s40294-016-0015-x, 2016.

[15] A. Dogan and D. Birant, "A Weighted Majority Voting Ensemble Approach for Classification," Proc. of the 4th IEEE Int. Conf. on Computer Science and Eng. (UBMK), pp. 1-6, Samsun, Turkey, 2019.

[16] R. Olfati-Saber, J. A. Fax and R. M. Murray, "Consensus and Cooperation in Networked Multi-agent Systems," Proceedings of the IEEE, vol. 95, no. 1, pp. 215-233, 2007.

[17] Y. Chen, J. Liu, H. Zhao and H. Su, "Social Structure Emergence: A Multi-agent Reinforcement

125

"Consensual Based Classification As Emergent Decisions in a Complex System", R. Mazouzi et al.

Learning Framework for Relationship Building," Proc. of the 19th International Conference on Autonomous Agents and Multi Agent Systems (AAMAS '20), pp. 1807-1809, 2020.

[18] C. Yang and L. Jiamou, "Social Capital Games As a Framework for Social Structural Pattern Emergence," Proc. of the 12th IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 309-316, The Hague, Netherlands, 2020.

[19] O. Paunovski, G. Eleftherakis and T. Cowling, "Disciplined Exploration of Emergence Using Multi-agent Simulation Framework," Computing and Informatics, vol. 28, pp. 369-391, 2009.

[20] O. Paunovski, G. Eleftherakis and A. J. Cowling, "A Fuzzy Reasoning Approach to Automated Detection of Emergent Herd Formations in Computer Aided Simulation," Proc. of the 3rd South East European Doctoral Student Conference, vol. 2, pp. 150-165, Thessaloniki, Greece, 2008.

[21] V. Robu, H. Halpin and H. Shepherd, "Emergence of Consensus and Shared Vocabularies in Collaborative Tagging Systems," ACM Transactions on the Web, vol. 3, no. 4, pp. 1-14, 2009.

[22] X. Wang, D.-W. Bi, L. Ding and S. Wang, "Agent Collaborative Target Localization and Classification in Wireless Sensor Networks," Sensors, vol. 8, pp. 1359-1386, 2007.

[23] R. Mazouzi, L. Seddiki, C. de Runz and H. Akdag, "Towards a Collective and Distributed System for Consensual Data Classification," HAL-EGC 2014, hal-03251347, version 1, 2014.

[24] L. Maystre, N. Kumarappan, J. Bütepage and M. Lalmas, "Collaborative Classification from Noisy Labels," Proc. of the 24th International Conference on Artificial Intelligence and Statistics, Proc. of Machine Learning Research, vol. 130, pp. 1639-1647, 2021.

[25] W. Sun, Z. Wang, G. Zhang and B. Dong, "MACCA: A SDN Based Collaborative Classification Algorithm for QoS Guaranteed Transmission on IoT," Proc. of the 15th International Conference on Advanced Data Mining and Applications (ADMA 2019), pp. 815-827, Springer, 2019.

[26] C. D. Stefano, A. D. Cioppa and A. Marcelli, "An Adaptive Weighted Majority Vote Rule for Combining Multiple Classifiers," Proc. of the IEEE International Conference on Pattern Recognition (ICPR), pp. 192-195, Quebec City, Canada, 2002.

[27] K. P. Murphy, "Naive Bayes Classifiers," University of British Columbia, vol. 18, no. 60, pp. 1-8, 2006.

[28] R. Hickey, "A Note on the Measurement of Randomness," Journal of Applied Probability, vol. 19, no. 1, pp. 229-232, 1982.

[29] J. Lin, "Divergence Measures Based on the Shannon Entropy," IEEE Transactions on Information Theory, vol. 37, no. 1, pp. 145-151, 1991.

[30] J. E. Hurtado and A. H. Barbat, "Monte Carlo Techniques in Computational Stochastic Mechanics," Archives of Computational Methods in Engineering, vol. 5, no. 1, pp. 3-29, 1998.

[31] M. Tavallaee. E. Bagheri, W. Lu and A. Ghorbani, "A Detailed Analysis of the KDD CUP 99 Data Set," Proc. of the IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA 2009), pp. 1-6, Ottawa, Canada, 2009.

[32] M. Lichman, "Connect-4", UCI Machine Learning Repository, 2013.

[33] J. A. Blackard, "Comparison of Neural Networks and Discriminant Analysis in Predicting Forest Cover Types," PhD Thesis, Computer Science Dept. Fort Collins, Colorado State University, USA, 1998.

ملخص البحث:

يتضمّن العمل الذي نعرضه في هذه الورقة التعامل مع الانبثاق في المصنِّفات المتعدّدة الضّخمة التي نعدُّها نُظماً معقّدة. فنحن نهدف الى بناء نظامٍ تعاوني لتصنيف البيانات على نحوٍ مُراقَب نتوقّع منهُ أن يقدّم أداءً أفضل مقارنةً بالمصنِّفات التّقليدية. وبنمذجة المصنِّف المتعدّد الضّخم كنظامٍ مُتعدّد العوامل، فإنّه يتكون من عددٍ كبيرٍ من العوامل المترابطة فيما بينها وفْق جِوارٍ معيّن؛ إذ يلعب كلّ عاملٍ دوْر مصنِّفٍ ضعيف. وعلى المستوى الميكروي، فإنّ التّفاعل الابتدائي بين العوامل يتمثّل في جمْع نتائج تصنيف كلٍّ منها. ويقوم كلّ عاملٍ وفقاً لقاعدة تصويت الأغلبية بجمع نتائجه مع نتائج جِوارِه من العوامل، آخذاً أداء كلٍّ منها بعيْن الاعتبار. وتتكرّر هذه العملية بشكلٍ متواصِل وبصورة دوريةٍ داخل جِوار كلّ عاملٍ. وبذلك تنشأ ديناميّة معقّدة داخل النّظام. وبعد زمنٍ معيّن، تستقر هذه الديناميّة المعقّدة، سامحةً بعرْض بنْيةٍ انبثاقيّة ستتمّ ملاحظتها على المستوى الماكروي وتُعدّ توقُّعاً رضائياً للأصناف بالنسبة للبيانات التي نرغبُ في تصنيفها. والجدير بالذكر أنّ النتائج التّجريبية التي حصلنا عليها والمقارنة مع نتائج المصنِّفات التّقليدية تُظهر إمكانيّة الطّريقة المقترحة في تحسين التّصنيف وأن تكون بديلاً لجمْع المصنِّفات ومراكمتها.

HYBRIDIZATION OF ARITHMETIC OPTIMIZATION WITH GREAT DELUGE ALGORITHMS FOR FEATURE SELECTION PROBLEMS IN MEDICAL DIAGNOSIS

Mohammed Alweshah

ABSTRACT

In the field of medicine, there is a need to filter data to find information that is relevant for specific research problems. However, in the realm of scientific study, the process of selecting the appropriate data or features is a substantial and challenging problem. Therefore, in this paper, two wrapper feature selection (FS) methods based on novel metaheuristic algorithms named the arithmetic optimization algorithm (AOA) and the great deluge algorithm (GDA) were used to attempt to tackle the medical diagnostics challenge. Two methods, AOA and AOA-GD were tested on 23 medical benchmark datasets. According to all of the experimental data, the hybridization of the GDA with the AOA considerably increased the AOA's search capability. The AOA-GD method was then compared with two previous wrapper FS approaches; namely, the coronavirus herd immunity optimizer with greedy crossover operator (CHIO-GC) and the binary moth flame optimization with Lévy flight (LBMFO_V3). When applied to the 23 medical benchmark datasets, the AOA-GD achieved an accuracy rate of 0.80, thereby surpassing both the CHIO-GC and LBMFO V3.

KEYWORDS

1. INTRODUCTION

In the last decade, artificial intelligence (AI) has undergone significant development and there are signs that it has already reached the level of being able to give genuine solutions to healthcare problems, heralding the dawn of a revolution in the field of medicine [1]. However, the use of AI raises some challenges, which mostly concern the extent of the ability of AI to simulate human skills, such as logical thinking. Moreover, it excels at analyzing huge data and reaching correct scientific findings throughout record durations. In recent years, a broad variety of AI initiatives have been proposed for gathering and analyzing massive amounts of health data, the most significant of which is machine learning (ML) [2]-[3].

Machine learning is now widely utilized for evaluating medical data and much work has been done in the area of medical diagnosis to address specific diagnostic issues [4]-[5]. In specialist hospitals or departments, data on proper diagnoses is frequently available in the form of medical records [3], [6]. All that is required is to enter the patient data with known proper diagnoses into a computer software program that then runs a learning process [7]-[9]. However, there is still a need to improve classifier performance, which has led to the usage of the feature selection (FS) approach as a means of simplifying the already available classifiers [10]-[12].

In data preprocessing, FS is an essential technique that is used to identify a sub-set of associated attributes. Feature selection is particularly important in supervised learning, because it optimizes a specific function to improve prediction accuracy by picking the relevant features in a given class label. For the goal of optimizing the prediction model, several FS approaches are utilized and have been developed [13]-[14].

To put it simply, the FS process finds and retains only those features that are the most relevant to the problem at hand [15]. By eliminating irrelevant as well as common characteristics, the FS method decreases the number of features that a classifier has to learn, which then reduces the training time and the number of features that the classifier has to evaluate, thereby resulting in improved classification performance [16]-[17]. The FS method is used to choose the optimal sub-set of features from the whole feature space in order to provide the necessary elucidations about the learning operations [18].

M. Alweshah is with Prince Abdullah Bin Ghazi Faculty of Information and Communication Technology, Al-Balqa Applied University, Al-Salt, Jordan. Email: weshah@bau.edu.jo

Feature selection consists of four major steps: starting up, searching, evaluating sub-sets of features and reaching the stopping condition [19]. To date, several FS approaches have been developed and utilized to attempt to optimize prediction models [20].

Researchers are also constantly attempting to find ways to enhance the accuracy of ML by utilizing another algorithm with a classifier algorithm in the learning model [21]-[23]. To express a single metaheuristic mechanism as an optimizer, one algorithm is used in conjunction with the learning model [24]. Many research studies have recently proven the efficacy of similar algorithms in achieving improved outcomes and in improving the method for picking a variety of characteristics [1]. Nature has inspired some of the most successful metaheuristics [25]. When tracing the search method, metaheuristics takes the information obtained into account [10]. Furthermore, it generates new ways of connecting a single, efficient approach or more [26].

In this paper, the arithmetic optimization algorithm (AOA), a novel metaheuristic algorithm created by Abualigah et al. in 2020 [27] for the field of medical diagnostics is utilized to address FS concerns. Multiplication, division, subtraction and addition are the arithmetic operators employed by the AOA; these operators reflect the standard calculation techniques used to investigate numbers. These basic operators are used as a mathematical optimization to choose the assessment that can help from a set of candidate replacements consisting of a set of criteria (solutions).

The AOA is employed in two techniques in this study to choose the most valuable and first used qualities in medical datasets in their basic form. Then, the AOA is hybridized with the great deluge algorithm (GDA) [28] in an attempt to increase its exploration capability. The hybrid approach is named the AOA-GD. The two suggested methods, AOA and AOA-GD are implemented in a wrapper model using a K-nearest neighbor (KNN) classifier and their performance is compared with those of other methods in the literature.

The rest of this paper is organized as follows: First, the AOA, the GDA and the proposed approaches for FS are discussed in Sections 2, 3 and 4, respectively. Then, the experiments and results are presented and discussed in Section 5. Finally, a conclusion is presented in Section 6.

2. THE ARITHMETIC OPTIMIZATION ALGORITHM

Arithmetics, along with geometry, algebra and analysis, are among the fundamental components of number theory and among the most significant aspects of modern mathematics. The conventional calculation methods that are used to analyze numbers are multiplication, division, subtraction and addition and are called arithmetic operators [29]. These basic operators have been utilized in mathematical optimization techniques to select the best element from a group of potential alternatives based on certain criteria (solutions) [30]. In problems with optimization, there are numerous quantitative domains, such as engineering, economics and computer science, as well as operations research and industry and the creation of improved solution methodologies has therefore long been of interest to mathematicians [31].

Regardless of the variations in the metaheuristic algorithms established for population-based optimization techniques, the optimization process is divided into two stages: exploration and exploitation. To prevent local solutions, the former includes the deployment of search agents that cover a large search field. The latter is the enhancement of the correctness of the obtained solutions during the exploration phase [32].The mathematical model is utilized to make a recommendation for the AOA [33].

2.1 Initialization Phase

The AOA optimization procedure begins with a set of randomly generated candidate solutions (X), as seen in the matrix below (Equation (1)) and the best candidate solution found in each iteration is deemed to be the best-obtained or nearly optimum solution so far.

Before starting its work, the AOA must decide on the search phase (i.e., exploration or exploitation). This is done by using the math optimizer accelerated (MOA) function, which is a coefficient that is calculated using Equation (2) and employed in the succeeding search phases:

$$X = \begin{cases} X1,1 & \dots & \dots & X1,J & X1,N-1 & X1,N \\ X2,1 & \dots & \dots & X2,J & X2,N-1 & X2,N \\ X3,1 & \dots & \dots & X3,J & X3,N-1 & X3,N \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ XN,1 & \dots & \dots & XN,J & XN,N-1 & XN,N \end{cases} \tag{1}$$

$$\text{MOA (C\_Iter)} = \text{Min} + \text{C\_Iter} \times \text{Max} - \left\{ \frac{\text{Min}}{\text{M\_Iter}} \right\} \tag{2}$$

MOA (C Iter) is the function value at the 't'th iteration as calculated by Eq. (2). The current iteration is the number of iterations between 1 and the maximum number of iterations (M Iter) (C Iter). The accelerated function's minimum and maximum values are presented in Min and Max, respectively.

2.2 Exploration Phase

The exploration operators of the AOA use two basic search methods (the D search strategy and the M search strategy), to randomly examine different regions of the search space in order to discover a better solution (3). This phase of searching is conditioned for the condition of r1 > MOA by the MOA function, where r1 is a random number.

The first operator (D) is conditioned by r2 0.5 in this phase, while the other operator (M) is disregarded until the former operator completes its task. If D is unable to complete its task, instead of D, the duty is handed to the second operator (M). (It should be noted that r2 is a random integer.). To replicate the behavior of arithmetic operators, the simplest rule is employed [20].

For the exploration portions, two position updating equations are provided in this paper. The following is the initial position update equation:

$$xi,j\,(C_{Iter} + 1) = \begin{cases} \text{best(xj)} \div (\text{MOP} + \varepsilon) \times ((UBj - LBj) \times M + LBj), & r2 < 0.5. \\ \text{best(xj)} \times \text{MOP} \times ((UBj - LBj) \times M + LBj), & \text{otherwise} \end{cases} \tag{3}$$

where xi (C_ Iter + 1) signifies the ith solution for next iteration, xi and j(C_Iter) signifies the jth position of the current iteration's ith solution, best(xj) and is the best solution achieved thus far. UBj and LBj denote the upper and lower limits of the jth position, respectively. In the experiments conducted for this study, M is a control parameter for altering the search process and is set at 0.5. The second position updating equation is as follows:

$$\text{MOP (C\_Iter)} = 1 - \left\{ \frac{\text{C\_Iter } 1/\alpha}{\text{M\_Iter } 1/\alpha} \right\} \tag{4}$$

where MOP (C_Iter) represents the function value at the 't' iteration, C_Iter indicates the current iteration, (M_Iter) denotes the maximum number of iterations and MOP(C_Iter) indicates the cost function at the 't'th iteration and MOP (C_Iter) signifies the function value at the 't'th iteration, which is set to 5 in the experiments conducted for the current work. Note that this is a crucial parameter that influences the accuracy of exploitation across repetitions.

2.3 Exploitation Phase

Subtraction (S) and addition (A) mathematical processes provide high-density outputs which represent the exploitation search process. However, unlike the other operators (D and M), these operators (S and A) may easily reach the objective owing to their low dispersion. As a result, the exploitation search might find a near-optimal solution that can be established after several attempts (iterations).

The MOA cost function for the constraint that r1 is not greater than the current MOA(C_ Iter) value (see Equation (4)) is conditioned for this phase of searching (exploitation search by executing S or A). As shown in Equation (5), the AOA exploitation operators (S and A) delve deeply into specified dense parts of the search space in order to find a better solution.

$$xi,j\,(C\_Iter + 1) = \begin{cases} \text{best(xj)} - \text{MOP} \times ((UBj - LBj) \times M + LBj), & r3 < 0. \\ \text{best(xj)} + \text{MOP} \times ((UBj - LBj) \times M + LBj), & \text{otherwise} \end{cases} \tag{5}$$

In this phase, which involves performing a deep search of the search space, the first operator (S) is conditioned by r30.5 (first rule in Equation (5)), while the other operator (A) is disregarded until the former operator (S) completes its task. If S is unable to complete its task, the second operator (A) is

utilized to accomplish the current task instead of S. The processes in this phase are identical to the partitioning in the previous phase. This method aids exploratory search strategies in finding the ideal answer while keeping a diversity of candidate solutions.
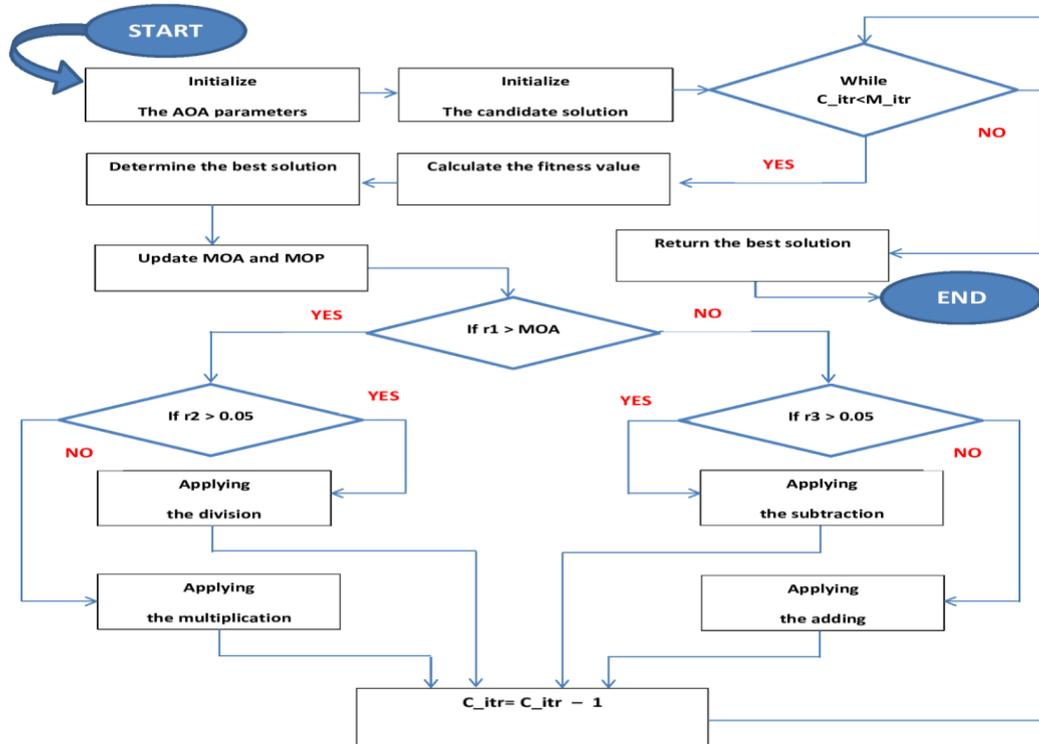

Figure 1. Flowchart of proposed AOA [20].

The M parameter is carefully chosen to create a random value at each iteration, which allows exploration to continue not only during the first but also during the last iteration. This element of the search process is highly effective when local optima stagnation occurs, particularly in the last iterations. The final location found can fall within a stochastic range specified by the search scope's locations of D, M, S and A. Various solutions update their locations stochastically about the near-optimal solution's region, whereas D, M, S and A estimate the position of the near-optimal solution in other ideas. All of these steps are clarified in the AOA flowchart illustrated in Figure 1.

All of these steps are clarified in Algorithm 1 below, which contains the pseudocode of the AOA:

Algorithm 1: The pseudo-code of the AOA.
1. Set the parameters for the AOA.
2. Randomize the positions of the solutions.
3. Determine the fitness value.
4. While C_itr<M_itr
5. Find the best solution.
6. Update the MOA value.
7. Update the MOP value.
8. Calculate the Fitness Function for the new solution.
9. for (i = 1 to Solution) do
10. if rand < 0.5 then
11. Create a set of random numbers between [0, 1].
12. if r1 > MOA then
13. if r2 > 0.5 then
14. Using the first rule in Equation (3), modify the positions of the ith solutions.
15. Else
16. Using the second rule in Equation (3), modify the positions of the ith solutions.
17. End if
18. Else

130

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 08, No. 02, June 2022.

19. if r3 > 0.5 then
20. Using the first rule in Equation (5), modify the positions of the ith solutions.
21. Else
22. Using the second rule in Equation (5), modify the positions of the ith solutions.
23. Enf if
24. End if
25. End if
26. End for
27. C_Iter = C_Iter + 1
28. End while
29. Return the best solution.

3. GREAT DELUGE ALGORITHM

The great deluge algorithm (GDA) was invented by Dueck in 1993 [34]. The GDA operates similarly to simulated annealing (SA), with the exception that the GDA uses an upper limit (commonly referred to as the water level) as the acceptability barrier rather than a temperature. The GD technique starts with a boundary equal to the original solution's quality. If the cost (objective value) is much less than the boundary, which is reduced at a predetermined rate in each iteration, it accepts lesser options (known as the decay rate). The GDA has just one parameter (the decay rate), which gives it an advantage over the SA, since the effectiveness of a metaheuristic method is dependent on parameter tweaking [35].

Additionally, in comparison to SA, the GD method is less dependent on parameters. In actuality, the GD has only two parameters: the quality of the solution and the amount of time it takes to compute the solution [36]. The best solution is always accepted by the GDA. The worst solution, on the other hand, might be retained if its quality is less than or equal to a defined upper limit (water level), which is implemented to cope with minimization concerns. The water level is used as the initial value of the solution objective function and it is iteratively increased by a constant upper pound (UP) while the algorithm runs [37]. Algorithm 2 shows the pseudo-code for the typical GD algorithm.

Algorithm 2: The pseudo-code for the typical GD algorithm.
1: Pick a good initial setting.
2: Pick "rain speed" UP > 0.
3: Get an initial WATER-LEVEL > 0.
4: Use a new setting, which is a stochastic tiny setting.
5: Updates to the old setup
6: Measure E:= new quality setting
7: If E WATER-LEVEL is true, then
8: Then the previous setting is equal to the new setting.
9: WATER-LEVEL:= WATER-LEVEL+UP
10: If there has been no improvement in quality for a long period or if there have been too many iterations
11: Stop
12: End if
13: Output

4. PROPOSED METHOD

Two models are suggested in the current work. Both models use a wrapper FS method to select the most significant features inside a dataset. The first is based on the basic AOA, while the second involves combining the AOA with the GDA to achieve a balance between exploration and exploitation in the AOA.

In more detail, the current study proposes a progressive hybridization of the AOA and GD. The GD is incorporated into the AOA improvement process during the hybridization process. The hybridization process begins with a certain number of repetitions of the AOA. After the specified number of iterations, the GDA receives the best solution and highest fitness identified so far by the AOA and starts its improvement process. The obvious solution and fitness that the GDA discovers are then submitted back to the AOA to continue the process of development. This reciprocal procedure continues until all of the AOA iterations have been completed and the stopping condition has been fulfilled.

The proposed AOA-GD approach uses AOA to produce the initial (solution) population of possible solutions in the AOA stage. The GDA calculates the fitness value of all candidates in the second stage to identify better solutions, ensuring efficient convergence, high-quality solutions and finally obtaining the ideal parameter values and therefore improving classification accuracy. Figure 2 depicts the suggested AOA-GD with KNN solution to the FS issue.

The current study investigates the accuracy of two techniques, AOA and AOA-GD, for the FS process in the area of medical diagnostics utilizing a wrapper FS methodology based on KNN. The KNN classifier was selected, since previous research has shown that it has a good classification efficiency when used to FS issues. Throughout the investigation, the number of closest neighbors (K) was fixed at five. The 5-NN approach was used to assess fitness during the training phase using internal N-fold cross-validation, with a total of five folds and the average error rate in the classification methodology was computed for each fold. The number of folds (N) and the number of nearest neighbors were determined using previous work (K).
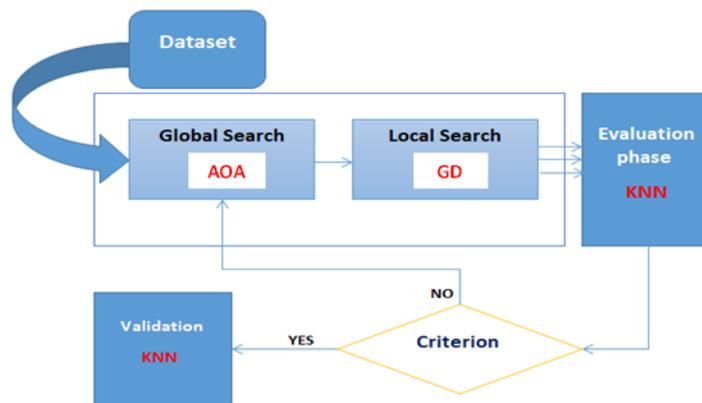

Figure 2. Proposed wrapper FS model based on AOA-GD method.

5. EXPERIMENTAL RESULTS AND DISCUSSION

This part explains the experimental design and the data processing procedures used to assess the suggested methodology's performance. Additionally, it compares the findings to those obtained using previous methodologies. The resulting instability is impacted by a variety of parameters, including accuracy, convergence rate and particular measures of central tilt. To guarantee a fair scientific investigation, same work settings and circumstances were employed across the trials. The program and its execution are powered by an Intel® CoreTM i7-6006U processor running at 2.00 GHz (four CPUs) and 2.0 GHz, with 8 GB of RAM. Matlab R2016a was used to create the model. Each dataset was partitioned into 70% for training and 30% for testing. The tests were run 30 times for each dataset, with each run consisting of 100 iterations.

5.1 Parameter Settings

The findings of some preliminary tests were used to specify the input parameters in the experiments, allowing the recommended technique to provide improved results. In order for the results to be consistent, the algorithm parameters were kept constant throughout the trial. Table 1 shows the parameter values utilized in each experiment.

Table 1. Parameter setups.

Parameter	Value
σ	0.1
щ	0.5
P	30
Max-itr	100
LB (lower bound)	0
UB (upper bound)	1

5.2 Datasets' Description

Medical data is defined as any information about an individual's health that is used to make normal patient care choices or to conduct diagnostic trials. Examples include administrative data, claims data, patient illness data and clinical trial data. The trial results were assessed using a collection of 23 medical benchmark datasets. The datasets were obtained from a number of sources, including UCI, KEEL and Kaggle, in addition to other well-known websites that provide FS medical datasets. Table 2 summarizes the properties of these datasets.

Table 2. Description of the datasets.

The names of datasets	The no. of instances	The no. of classes	The no. of features	The source
1. Diagnostic	569	2	30	Source:UCI
2. Original	699	2	9	Source:UCI
3. Prognostic	194	2	33	Source:UCI
4. Coimbra	115	2	9	Source:UCI
5. BreastEW	596	2	30	Source:UCI
6. Retinopathy	1151	2	19	Source:UCI
7. Dermatology	366	6	34	Source:UCI
8. ILPD-Liver	583	2	10	Source:UCI
9. Lymphography	148	4	18	Source:UCI
10. Parkinsons	194	2	22	Source:UCI
11. ParkinsonC	755	2	753	Source:UCI
12. SPECT	267	2	22	Source:KEEL
13. Cleveland	297	5	13	Source:KEEL
14. HeartEW	270	2	13	Source:KEEL
15. Hepatitis	79	2	18	Source:KEEL
16. SAHear	461	2	9	Source:KEEL
17. Spectfheart	266	2	43	Source:KEEL
18. Thyroid0387	7200	3	21	Source:KEEL
19. Heart	302	5	13	Source:Kaggle
20. Pima-diabetes	768	2	9	Source:Kaggle
21. Leukemia dataset	72	2	7129	The source: https: //jundongl.github.io/scikit-feature/ datasets.html
22. Colon dataset	62	2	2000	The source: https: //jundongl.github.io/scikit-feature/ datasets.html
23. ProstateGE dataset	102	2	5966	The s ource :https: //jundongl.github.io/scikit-feature/ datasets.html

As shown in Table 2, the 23 datasets cover a range of case studies on medical diagnosis with varying architectures. Hence, the efficacy of the AOA and AOA-GD was determined by testing them on various problems with varying features. 70% of the datasets were utilized for training purposes and 30% for testing. Each dataset was tested 30 times, with each run consisting of 100 iterations.

5.3 Experimental Result

The recall, accuracy, precision, F-measure, error rate, number of features selected and convergence speed of the two proposed approaches, AOA and AOA-GD, were evaluated. Table 3 compares the accuracy rate and selection size of the two approaches after 30 runs on each dataset.

Table 3. AOA and AOA-GD accuracy and selection size results.

	The dataset name	Accuracy		Selection Size	
		(AOA)	(AOA-GD)	(AOA)	(AOA-GD)
1	Diagnostic	.8470	.9097	15.381	12.9140
2	Original	.9300	.9762	7.0005	5.6201
3	Prognostic	.5888	.6501	17.9302	13.9271
4	Coimbra	.8105	.9100	4.0081	3.3810
5	BreastEW	.9017	.9531	14.9917	14.0027
6	Retinopathy	.5047	.6594	8.0590	6.7015
7	Dermatology	.7101	.8206	19.5103	17.0099
8	ILPD-Liver	.6273	.7619	4.9140	4.0000

"Hybridization of Arithmetic Optimization with Great Deluge Algorithms for Feature Selection Problems in Medical Diagnosis", M. Alweshah.

9	Lymphography	.7593	.8511	11.4291	9.6910
10	Parkinsons	.6482	.7833	11.0610	10.7103
11	ParkinsonC	.6995	.8391	370.1920	360.0081
12	SPECT	.6109	.7108	10.0003	9.7201
13	Cleveland	.4792	.6163	8.1935	6.0003
14	HeartEW	.8741	.9207	6.0080	6.0619
15	Hepatitis	.6592	.7888	9.9104	9.0996
16	GDHear	.6208	.7259	5.0017	3.5039
17	Spectfheart	.6891	.7419	21.2710	19.6914
18	Thyroid0387	.9000	.9607	10.0041	7.0071
19	Heart	.7194	.8192	8.9914	7.1900
20	Pima-diabetes	.7005	.8201	5.8099	6.2814
21	Leukemia	.9819	.9914	3599.0092	3559.6091
22	Colon	.6430	.7228	1010.8192	989.2715
23	Prostate_GE	.4917	.6307	3038.0197	2961.0041

As seen in Table 3, the AOA-GD method bested the AOA in the whole datasets in terms of accuracy. This implies that if the search process of the AOA is changed, it is capable of producing more trustworthy findings.

As regards the number of features selected (FS size), the AOA-GD outperformed the AOA in 22 out of 24 datasets. Only in the Pima-diabetes and HeartEW datasets, the AOA able to outperform the AOA-GD. These findings illustrate the effectiveness of the AOA-GD modification in improving the exploratory search capacity of the AOA that enables it to discover the best basic solutions.

To further evaluate the findings and the classifier's capacity to provide trustworthy, correlated and similar solutions in all sequences for each dataset, the precision, recall and F-measure values were determined. Precision is the ratio of properly identified true positive IDs, recall is the ratio of correctly detected true positive IDs and the F-measure is the balance of the recall and precision ratios. Precision, recall and F-measure were computed as in [38]. Table 4 demonstrates how the AOA and AOA-GD methodologies' efficiency has been changed and focused over wholly datasets utilized in the experiment.

To compare the effectiveness of the AOA and AOA-GD approaches, a T-test was used. The presented

Table 4. Classification results for precision, recall and F-measure using AOA and AOA-GD.

	Datasets	Precision Result		Recall Result		F- Measure Result	
		AOA	AOA-GD	AOA	AOA-GD	AOA	AOA-GD
1	-Diagnostic	.950	.970	.960	.970	.960	.960
2	-Original	.930	.950	.950	.980	.930	.940
3	-Prognostic	.930	.960	.950	.980	.930	.940
4	-Coimbra	.880	.900	.940	.000	.660	.670
5	-BreastEW	.750	.790	.780	.750	.750	.760
6	-Retinopathy	.800	.860	.700	.850	.830	.820
7	-Dermatology	.870	.960	.880	.960	.870	.830
8	-ILPD-Liver	.920	.960	.930	.980	.960	.920
9	-Lymphography	.810	.900	.790	.890	.790	.710
10	-Parkinsons	.890	1.000	.880	1.000	.860	.930
11	-ParkinsonC	.820	.950	.830	.940	.630	.700
12	-SPECT	.780	.890	.870	.970	.740	.730
13	-Cleveland	.810	.900	.790	.880	.800	.780
14	-HeartEW	.740	.790	.750	.790	.730	.710
15	-Hepatitis	.930	.980	.930	.970	.920	.890
16	-GDHear	.780	.770	.720	.760	.770	.770
17	-Spectfheart	.950	.970	.960	.970	.960	.960
18	-Thyroid0387	.930	.950	.950	.980	.930	.940
19	-Heart	.930	.960	.950	.980	.930	.940
20	-Pima-diabetes	.880	.900	.940	1.000	.660	.670
21	-Leukemia	.750	.790	.780	.750	.750	.760
22	-Colon	.800	.860	.700	.850	.830	.820
23	-Prostate_GE	.870	.960	.880	.960	.870	.830

methodologies are used to calculate the findings statistics depending on the accuracy of the findings point to each dataset. Table 5 demonstrates the results of T-test along with a 95% confidence interval (the alpha value = 0.05) as well as the p-values and classification accuracy produced by the AOA and AOA-GD.

Table 5. T-test results for AOA and AOA-GD.

	Datasets	The Method Used	The Mean Value	The Std. Deviation Value	The Std. Error Mean Value	P-value Result
1-	-Diagnostic	AOA	0.8540	0.02673	0.00.488	00.00
		AOA-GD	0.9033	0.04046	0.00739	
2-	-Original	AOA	0.9233	0.02523	0.00461	00.00
		AOA-GD	0.9710	0.01062	0.00194	
3-	-Prognostic	AOA	0.5293	0.04638	0.00847	00.00
		AOA-GD	0.6716	0.04442	0.00811	
4-	-Coimbra	AOA	0.8006	0.04548	0.00830	00.00
		AOA-GD	0.8896	0.00999	0.00182	
5-	-BreastEW	AOA	0.8993	0.02100	0.00383	00.00
		AOA-GD	0.9400	0.01912	0.00349	
6-	-Retinopathy	AOA	0.4660	0.06106	0.01115	00.00
		AOA-GD	0.6436	0.02553	0.00466	
7-	-Dermatology	AOA	0.6690	0.04088	0.00746	00.00
		AOA-GD	0.8006	0.04548	0.00830	
8-	-ILPD-Liver	AOA	0.6423	0.02609	0.00476	00.00
		AOA-GD	0.7716	0.01744	0.00318	
9-	-Lymphography	AOA	0.7606	0.03483	0.00636	00.00
		AOA-GD	0.8343	0.02661	0.00486	
10-	-Parkinsons	AOA	0.6690	0.04088	0.00746	00.00
		AOA-GD	0.7903	0.01903	0.00347	
11-	-ParkinsonC	AOA	0.6856	0.06611	0.01207	00.00
		AOA-GD	0.8400	0.02197	0.00401	
12-	-SPECT	AOA	0.6073	0.02840	0.00518	00.00
		AOA-GD	0.6960	0.06667	0.01217	
13-	-Cleveland	AOA	0.4896	0.04173	0.00762	00.00
		AOA-GD	0.5966	0.02496	0.00456	
14-	-HeartEW	AOA	0.8540	0.02673	0.00488	00.00
		AOA-GD	0.9116	0.01783	0.00325	
15-	-Hepatitis	AOA	0.6690	0.04088	0.00746	00.00
		AOA-GD	0.7903	0.01903	0.00347	00.00
16-	-SAHear	AOA	0.6420	0.03089	0.00564	00.00
		AOA-GD	0.7036	0.01066	0.00195	
17-	-Spectfheart	AOA	0.6716	0.04442	0.00811	00.00
		AOA-GD	0.7303	0.03178	0.00580	
18-	-Thyroid0387	AOA	0.8986	0.04455	0.00813	00.00
		AOA-GD	0.9603	0.01474	0.00269	
19-	-Heart	AOA	0.7316	0.04009	0.00732	00.00
		AOA-GD	0.8126	0.02612	0.00477	
20-	-Pima-diabetes	AOA	0.7153	0.05557	0.01015	00.00
		AOA-GD	0.7956	0.03757	0.00686	
21-	-Leukemia	AOA	0.9876	0.01736	0.00317	00.00
		AOA-GD	0.9900	0.01017	0.00186	
22-	-Colon	AOA	0.6203	0.03634	0.00663	00.00
		AOA-GD	0.7176	0.05380	0.00982	
23-	-Prostate_GE	AOA	0.4750	63290.0	11550.0	00.00
		AOA-GD	0.6010	25370.0	04630.0	
24-	-Covid-19 dataset	AOA	0.9135	0.02523	0.00461	00.00
		AOA-GD	0.9370	0.01912	0.00349	

Table 5 demonstrates that the AOA-GD is more efficient than the original AOA, where the p-values in all datasets are below 0.0001. This data indicates that the AOA-GD is useful for resolving FS issues. It is generally understood that a consistent and fast ratio of convergence leads to superior solutions. Therefore, to further evaluate the efficiency of the proposed AOA and AOA-GD methods, their

"Hybridization of Arithmetic Optimization with Great Deluge Algorithms for Feature Selection Problems in Medical Diagnosis", M. Alweshah.

convergence speed behavior curves were studied in detail. Figure 3 shows the convergence speeds of the two proposed methods when employed to each of the 24 datasets over 30 runs.
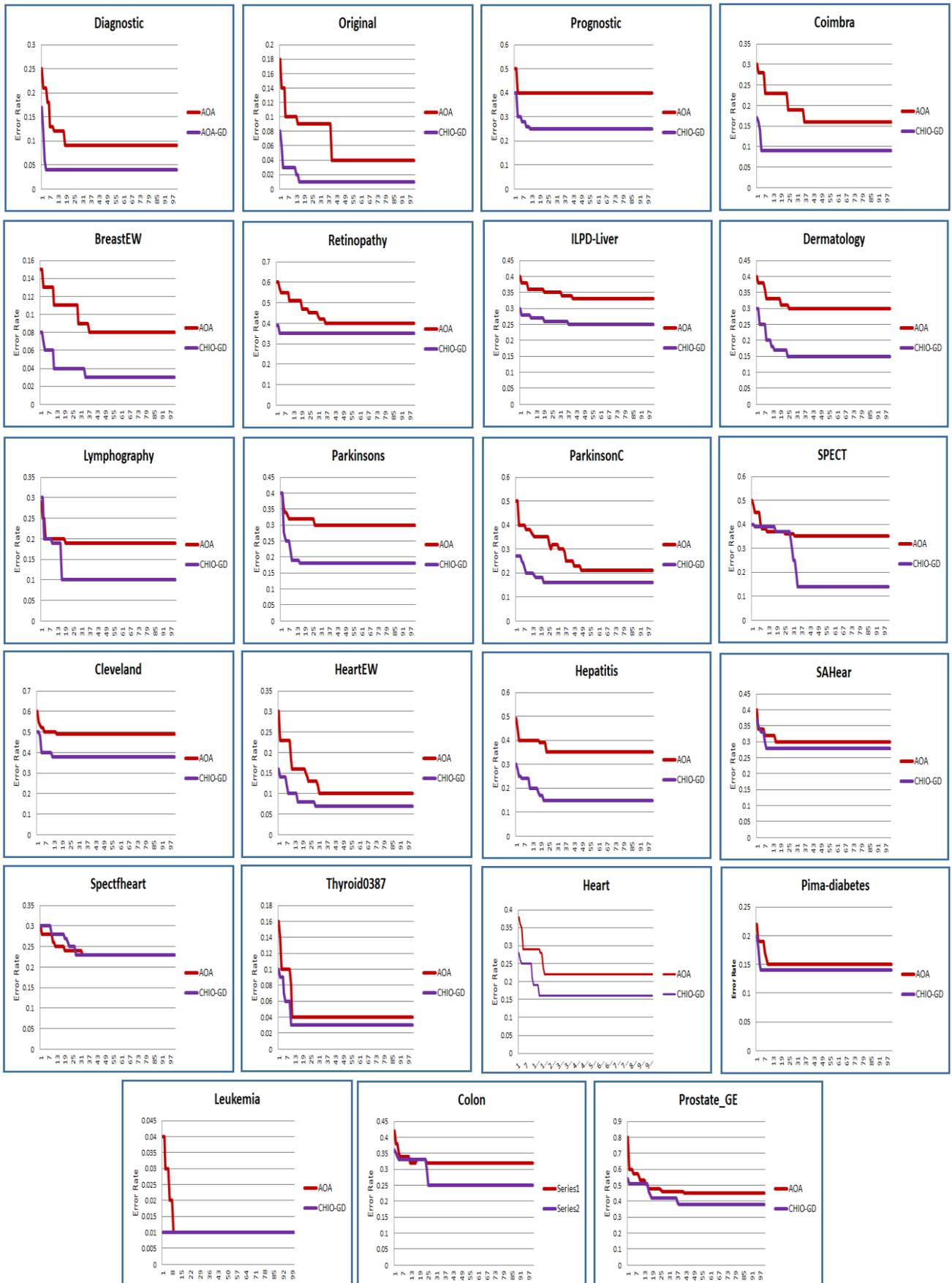

Figure 3. Convergence speeds of AOA and AOA-GD.

136

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 08, No. 02, June 2022.

The findings in Figure 3 show that the AOA-GD was able to increase classification accuracy at a faster convergence time than the AOA. This was achieved by the GDA boosting the global search capacity of the original AOA.

5.4 Comparison with Previous Methods

The better of the two suggested approaches, AOA-GD, was compared to the CHIO-GC [39] (M1) and LBMFO-V3 (M2) [40] on the 23 medical datasets to examine the dependability of the proposed algorithm and its capacity to create a high degree of classification accuracy while decreasing the number of chosen characteristics. The classification accuracy and FS size of the AOA-GD were associated with those of the LBMFO-V3 and the CHIO-GC utilizing 23 medical datasets. The classification accuracy and FS size are calculated as average values over 30 runs. Table 6 displays the results.

Table 6. Comparison of the AOA-GD, CHIO-GC and LBMFO-V3 on the 23 medical benchmark datasets.

	Datasets	Average of Accuracy			Selection Size		
		AOA-GD	M1	M2	AOA-GD	M1	M2
1	Diagnostic	0.9097	0.9033	0.9100	12.9140	13.3700	13.9991
2	Original	0.9762	0.9710	0.9683	5.6201	5.1040	5.5000
3	Prognostic	0.6501	0.6716	0.5851	13.9271	14.6202	15.0126
4	Coimbra	0.9100	0.8896	0.9312	3.3810	3.6007	3.5103
5	BreastEW	0.9531	0.9400	0.9398	14.0027	13.7303	13.9714
6	Retinopathy	0.6594	0.6436	0.5380	6.7015	.72647	6.9002
7	Dermatology	0.8206	0.8006	0.8442	17.0099	18.4900	18.3541
8	ILPD-Liver	0.7619	0.7716	0.7143	4.0000	4.0000	4.0000
9	Lymphography	0.8511	0.8343	0.8002	9.6910	.100622	9.7520
10	Parkinsons	0.7833	0.7903	0.7689	10.7103	9.7383	10.3584
11	ParkinsonC	0.8391	0.8400	0.8190	360.0081	365.8322	369.1070
12	SPECT	0.7108	0.6960	0.6576	9.7201	9.6050	10.7832
13	Cleveland	0.6163	0.5966	0.5333	6.0003	6.8097	6.6899
14	HeartEW	0.9207	0.9116	0.9388	6.0619	7.0105	6.3100
15	Hepatitis	0.7888	0.7903	0.7500	9.0996	8.2011	8.3569
16	SAHear	0.7259	0.7036	0.6992	3.5039	.31551	3.2222
17	Spectfheart	0.7419	0.7303	0.7013	19.6914	21.0030	20.4598
18	Thyroid0387	0.9607	0.9603	0.9776	7.0071	8.0116	8.4563
19	Heart	0.8192	0.8126	0.7603	7.1900	6.1505	6.2752
20	Pima-diabetes	0.8201	0.7956	0.8065	6.2814	.68387	6.7612
21	Leukemia	0.9914	0.9900	1.0000	3559.6091	3560.5107	3570.7137
22	Colon	0.7228	0.7176	0.6667	989.2715	1000.0067	991.5551
23	Prostate_GE	0.6307	0.6010	0.5056	2961.0041	2979.4116	2984.7153
	Average	0.8071	0.7983	0.7746	349.6698	351.4142	351.9462

In 18 datasets, including Diagnostic original, Coimbra, BreastEW, Retinopathy, Dermatology, Lymphography, SPECT, Cleveland, HeartEW, SAHear, Spectfheart, Thyroid0387, Heart, Pima-diabetes, Leukemia, Colon and Prostate GE, the AOA-GD method outperformed the CHIO-GC method with regards to classification accuracy. Additionally, the AOA-GD approach outperformed the LBMFO-V3 technique in 17 datasets, including the Original, Prognost, BreastEW, Retinopathy, Lymphography SPECT, Parkinsons, ILPD-Liver, Colon, Cleveland, Hepatitis, SAHear, Spectfheart,,Heart, Parkinson C, Pima-diabetes and Prostate GE. The AOA-GD strategy outperformed the CHIO-GC and LBMFO-V3 approaches across each dataset, with on average accuracy of (0.8071).

In 16 datasets, the AOA-GD technique outperformed the CHIO-GC method in terms of selection size, including Diagnostic, Prognostic, Coimbra, Retinopathy, Dermatology, ILPD-Liver, Lymphography, ParkinsonC, Cleveland, HeartEW, Spectfheart, Thyroid0387, Pima-diabetes, Leukemia, Colon and Prostate GE. Additionally, the AOA-GD technique outperformed the LBMFO-V3 method in 17 datasets, including Diagnostic, Prognostic, Coimbra, Retinopathy, Dermatology, ILPD-Liver, Lymphography, ParkinsonC, SPECT, Cleveland, HeartEW, Spectfheart, Thyroid0387, Pima-diabetes,

Leukemia, Colon and Prostate GE. Across all datasets, the AOA-GD had an average selection size of 349.6698 features. Figure 4 graphically illustrates, respectively, the average accuracy and average selection size of the three methods.
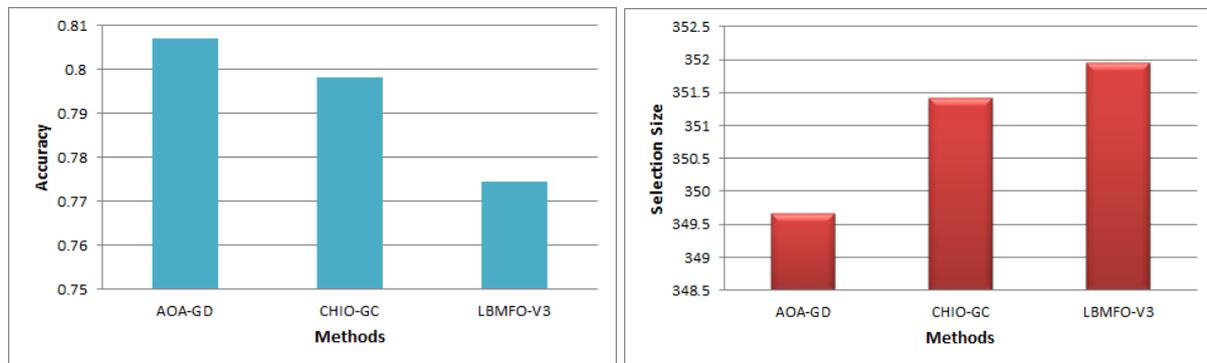

Figure 4. Average accuracy and average selection size of all three methods on the 23 medical benchmark datasets.

5.5 Discussion

The initial results for the AOA demonstrated that the search mechanism of the algorithm finds a reasonable balance between exploration and exploitation [41]. The capacity of metaheuristic algorithms to uncover optimum solutions throughout the search process is one of its most important characteristics. According to the findings, the suggested hybridization of the GDA with the AOA improves the AOA's exploration capabilities, allowing it to better choose the initial characteristics necessary to fulfill the aim of optimizing the solution. In other words, the suggested change aided in improving the balance between exploration and exploitation.

In all 23 datasets, Table 4 demonstrates that the AOA-GD approach surpassed the AOA approach in terms of classification accuracy, highlighting the utility of the proposed hybrid AOA-GD technique for striking an acceptable balance between exploration and exploitation. Additionally, as shown in Table 4, Figure 6 and Figure 7, the AOA-GD excelled in terms of the highest and lowest accuracy achievable throughout each run. Finally, the AOA-GD demonstrated to be beneficial in narrowing the accuracy gap between maximum and minimum values and accelerating convergence.

6. CONCLUSION

The FS problem is one of the most pressing concerns for academics across a broad range of disciplines; metaheuristics has been widely used in recent years in feature services (FS) to reduce the number of features mandatory to obtain satisfactorily honest results, with the objective of growing to have reliability and performance. The AOA metaheuristics was used in the current work for two proposed models for addressing FS issues in medical diagnosis. The first model was based on the basic AOA, while the second entailed integrating the AOA with the GDA in order to obtain a better balance between exploration and exploitation in the AOA. 23 medical datasets were used to evaluate the suggested approaches. Many measures were used to compare the two strategies, including number of selected features, classification accuracy, recall, precision, convergence speed, F-measure and T-test. The findings of all data analyses indicated that the AOA-GD enhanced the exploratory capabilities of the original AOA. Moreover, the feature selection size and classification accuracy of the AOA-GD were compared to those of other approaches previously published. The investigation's findings indicated that the AOA-GD technique outperformed the CHIO-GC and LBMFO-V3 wrapper approaches. AOA-GD surpassed the CHIO-GC and LBMFO-V3 in the majority of medical benchmark datasets with an accuracy rate of 0.80 and a selection size rate of 349 features, according to the results of the study. These encouraging findings were obtained as a consequence of a well-balanced AOA-GD search phase during the identification of appropriate solutions, which increased the pace of convergence. This optimum balance was reached by combining the AOA with the GDA, since the GDA was able to correct the unacceptable solutions that were obtained during premature convergence and while confined in a narrow optimal search space, respectively.

ACKNOWLEDGEMENTS

CONFLICT OF INTEREST

The authors certify that they have no affiliations with or involvement in any organization or entity with any financial or non-financial interest in the subject matter or materials discussed in this manuscript.

REFERENCES

[1]   M. Alweshah, S. A. Khalaileh, B. B. Gupta, A. Almomani, A. I. Hammouri and M. A. Al-Betar, "The Monarch Butterfly Optimization Algorithm for Solving Feature Selection Problems," Neural Computing and Applications, vol. 2020, pp. 1-15, 2020.

[2]   C. Lam, A. Siefkas, N. S. Zelin et al., "Machine Learning As a Precision-medicine Approach to Prescribing COVID-19 Pharmacotherapy with Remdesivir or Corticosteroids," Clinical Therapeutics, vol. 43, no. 5, pp. 871-885, 2021.

[3]   M. Alweshah, M. Al-Sendah, O. M. Dorgham, A. Al-Momani and S. Tedmori, "Improved Water Cycle Algorithm with Probabilistic Neural Network to Solve Classification Problems," Cluster Computing, vol. 23, pp. 2703-2718, 2020.

[4]   S. Uddin, A. Khan, M. E. Hossain and M. A. Moni, "Comparing Different Supervised Machine Learning Algorithms for Disease Prediction," BMC Medical Informatics and Decision Making, vol. 19, pp. 1-16, 2019.

[5]   M. Alweshah, W. A. AlZoubi and A. Alarabeyyat, "Cluster Based Data Reduction Method for Transaction Datasets," Proc. of IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE), pp. 78-83, Langkawi, Malaysia, 2015.

[6]   K. Zukotynski, V. Gaudet, C. F. Uribe et al., "Machine Learning in Nuclear Medicine: Part 2—Neural Networks and Clinical Aspects," Journal of Nuclear Medicine, vol. 62, pp. 22-29, 2021.

[7]   N. Wang, Y. Huang, H. Liu, Z. Zhang, L. Wei, X. Fei and H. Chen, "Study on the Semi-supervised Learning-based Patient Similarity from Heterogeneous Electronic Medical Records," BMC Medical Informatics and Decision Making, vol. 21, pp. 1-13, 2021.

[8]   E. H. Houssein, I. E. Ibrahim, N. Neggaz, M. Hassaballah and Y. M. Wazery, "An Efficient ECG Arrhythmia Classification Method Based on Manta Ray Foraging Optimization," Expert Systems with Applications, vol. 181, p. 115131, 2021.

[9]   E. H. Houssein, M. E. Hosney and D. Oliva, "A Hybrid Seagull Optimization Algorithm for Chemical Descriptors Classification," Proc. of the International Mobile, Intelligent and Ubiquitous Computing Conference (MIUCC), pp. 1-6, Cairo, Egypt, 2021.

[10]   A. Almomani, M. Alweshah, S. Al Khalayleh, M. Al-Refai and R. Qashi, "Metaheuristic Algorithms-based Feature Selection Approach for Intrusion Detection," Chapter in Book: Machine Learning for Computer and Cyber Security, 1st Edn., CRC Press, pp. 184-208, 2019.

[11]   H. Al Nsour, M. Alweshah, A. I. Hammouri, H. Al Ofeishat and S. Mirjalili, "A Hybrid Grey Wolf Optimiser Algorithm for Solving Time Series Classification Problems," Journal of Intelligent Systems, vol. 29, no. 1, pp. 846-857, 2018.

[12]   M. Alweshah, L. Rababa, M. H. Ryalat, A. Al Momani and M. F. Ababneh, "African Buffalo Algorithm: Training the Probabilistic Neural Network to Solve Classification Problems," Journal of King Saud University-Computer and Information Sciences, DOI: 10.1016/j.jksuci.2020.07.004, 2020.

[13]   O. Dorgham, M. Alweshah, M. Ryalat, J. Alshaer, M. Khader and S. Alkhalaileh, "Monarch Butterfly Optimization Algorithm for Computed Tomography Image Segmentation," Multimedia Tools and Applications, pp. 1-34, DOI: 10.1007/s11042-020-10147-6, 2021.

[14]   M. Alweshah, "Solving Feature Selection Problems by Combining Mutation and Crossover Operations with the Monarch Butterfly Optimization Algorithm," Applied Intelligence, vol. 51, no. 1, pp. 1-24, 2020.

[15]   M. Alweshah, H. Rashaideh, A. I. Hammouri, H. Tayyeb and M. Ababneh, "Solving Time Series Classification Problems Using Support Vector Machine and Neural Network," International Journal of Data Analysis Techniques and Strategies, vol. 9, pp. 237-247, 2017.

[16]   B. Sathiyabhama et al., "A Novel Feature Selection Framework Based on Grey Wolf Optimizer for Mammogram Image Analysis," Neural Computing and Applications, vol. 33, pp. 14583–14602, 2021.

[17]   M. Alweshah, E. Ramadan, M. H. Ryalat, M. Almi'ani and A. I. Hammouri, "Water Evaporation Algorithm with Probabilistic Neural Network for Solving Classification Problems," Jordanian Journal of Computers and Information Technology (JJCIT), vol. 6, no. 1, pp. 1-15, 2020.

"Hybridization of Arithmetic Optimization with Great Deluge Algorithms for Feature Selection Problems in Medical Diagnosis", M. Alweshah.

[18]   M. Alweshah, O. A. Alzubi, J. A. Alzubi and S. Alaqeel, "Solving Attribute Reduction Problem Using Wrapper Genetic Programming," International Journal of Computer Science and Network Security (IJCSNS), vol. 16, p. 77, 2016.

[19]   M. Alweshah, S. Alkhalaileh, D. Albashish, M. Mafarja, Q. Bsoul and O. Dorgham, "A Hybrid Mine Blast Algorithm for Feature Selection Problems," Soft Computing, vol. 25, pp. 517-534, 2021.

[20]   H. Sun, J. Jin, R. Xu and A. Cichocki, "Feature Selection Combining Filter and Wrapper Methods for Motor-Imagery Based Brain–Computer Interfaces," International Journal of Neural Systems, vol. 31, no. 9, p. 2150040, 2021.

[21]   N. Darabi, A. Rezai and S. S. F. Hamidpour, "Breast Cancer Detection Using RSFS-based Feature Selection Algorithms in Thermal Images," Biomedical Engineering: Applications, Basis and Communications, vol. 33, no. 3, p. 2150020, 2021.

[22]   M. Salimian, A. Rezai, S. Hamidpour and F. Khajeh-Khalili, "Effective Features in Thermal Images for Breast Cancer Detection," Proc. of the 2nd National Conf. on New Technologies in Electrical and Computer Engineering, Isfahan, Iran, pp. 1-7, 2019.

[23]   M. Zarei, A. Rezai and S. S. Falahieh Hamidpour, "Breast Cancer Segmentation Based on Modified Gaussian Mean Shift Algorithm for Infrared Thermal Images," Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, vol. 9, no. 6, pp. 1-7, 2021.

[24]   H. Y. Chong, H. J. Yap, S. C. Tan et al., "Advances of Metaheuristic Algorithms in Training Neural Networks for Industrial Applications," Soft Computing, vol. 25, pp. 11209–11233, 2021.

[25]   Y. Xue, Y. Tang, X. Xu, J. Liang and F. Neri, "Multi-objective Feature Selection with Missing Data in Classification," IEEE Transactions on Emerging Topics in Computational Intelligence, DOI: 10.1109/TETCI.2021.3074147, pp. 1-10, 2021.

[26]   F. A. Hashim, E. H. Houssein, K. Hussain, M. S. Mabrouk and W. Al-Atabany, "Honey Badger Algorithm: New Metaheuristic Algorithm for Solving Optimization Problems," Mathematics and Computers in Simulation, vol. 192, pp. 84-110, 2022.

[27]   L. Abualigah, A. Diabat, S. Mirjalili, M. Abd Elaziz and A. H. Gandomi, "The Arithmetic Optimization Algorithm," Computer Methods in Applied Mechanics and Engineering, vol. 376, p. 113609, 2021.

[28]   D. Simon, "Biogeography-based Optimization," IEEE Transactions on Evolutionary Computation, vol. 12, pp. 702-713, 2008.

[29]   S. Khatir, S. Tiachacht, C. Le Thanh, E. Ghandourah, S. Mirjalili and M. A. Wahab, "An Improved Artificial Neural Network Using Arithmetic Optimization Algorithm for Damage Assessment in FGM Composite Plates," Composite Structures, vol. 273, p. 114287, 2021.

[30]   L. Abualigah, A. Diabat, P. Sumari and A. H. Gandomi, "A Novel Evolutionary Arithmetic Optimization Algorithm for Multilevel Thresholding Segmentation of Covid-19 ct Images," Processes, vol. 9, p. 1155, 2021.

[31]   M. Premkumar, P. Jangir, B. S. Kumar, R. Sowmya, H. H. Alhelou, L. Abualigah, A. R. Yildiz and S. Mirjalili, "A New Arithmetic Optimization Algorithm for Solving Real-world Multiobjective CEC-2021 Constrained Optimization Problems: Diversity Analysis and Validations," IEEE Access, vol. 9, pp. 84263-84295, 2021.

[32]   Y. Wu, "A Survey on Population-based Meta-heuristic Algorithms for Motion Planning of Aircraft," Swarm and Evolutionary Computation, vol. 62, p. 100844, 2021.

[33]   D. Gürses, S. Bureerat, S. M. Sait and A. R. Yıldız, "Comparison of the Arithmetic Optimization Algorithm, the Slime Mold Optimization Algorithm, the Marine Predators Algorithm, the Salp Swarm Algorithm for Real-world Engineering Applications," Materials Testing, vol. 63, pp. 448-452, 2021.

[34]   G. Dueck, "New Optimization Heuristics: The Great Deluge Algorithm and the Record-to-record Travel," Journal of Computational Physics, vol. 104, pp. 86-92, 1993.

[35]   D. Landa-Silva and J. H. Obit, "Great Deluge with Non-linear Decay Rate for Solving Course Timetabling Problems," Proc. of the 4th International IEEE Conference on Intelligent Systems, pp. 8-11-8-18, Varna, Bulgaria, 2008.

[36]   S. Kifah and S. Abdullah, "An Adaptive Non-linear Great Deluge Algorithm for the Patient-admission Problem," Information Sciences, vol. 295, pp. 573-585, 2015.

[37]   M. Mohmad Kahar and G. Kendall, "A Great Deluge Algorithm for a Real-world Examination Timetabling Problem," Journal of the Operational Research Society, vol. 66, pp. 116-133, 2015.

[38]   S. Kassaymeh, S. Abdullah, M. A. Al-Betar and M. Alweshah, "Salp Swarm Optimizer for Modeling the Software Fault Prediction Problem," Journal of King Saud University-Computer and Information Sciences, In Press, DOI: 10.1016/j.jksuci.2021.01.015, 2021.

[39]   M. Alweshah, S. Alkhalaileh, M. A. Al-Betar and A. A. Bakar, "Coronavirus Herd Immunity Optimizer with Greedy Crossover for Feature Selection in Medical Diagnosis," Knowledge-based Systems, vol. 235, p. 107629, DOI: 10.1016/j.knosys.2021.107629, 2021.

[40]   R. Abu Khurmaa, I. Aljarah and A. Sharieh, "An Intelligent Feature Selection Approach Based on Moth Flame Optimization for Medical Diagnosis," Neural Computing and Applications, vol. 33, pp. 7165-

7204, 2021.

[41] F. A. Hashim, K. Hussain, E. H. Houssein, M. S. Mabrouk and W. Al-Atabany, "Archimedes Optimization Algorithm: A New Metaheuristic Algorithm for Solving Optimization Problems," Applied Intelligence, vol. 51, pp. 1531-1551, 2021.

ملخص البحث:

فــي حقْــل الطِّـــبّ، ثمّـــة حاجـــةٌ الــى تصــفية البيانـــات لإيجـــاد المعلومـــات ذات العلاقـــة بمســائل بحثيـــة معينـــة. ومـــع ذلـــك، فــي إطــار الدّراســـة العلميـــة، فــإنّ عمليـــة انتقـــاء البيانـــات أو السِّمات الملائمة تبقى مسألةً أساسيةً تنطوي على الكثير من التّحديات.

لــذا، نقــدّم فــي هــذه الورقـــة طــريقتين مــن طُــرق انتقــاء السِّــمات (FS) بنـــاءً علــى خوارزميـــات عاليـــة التّوجيـــه، وبالـــذّات خوارزميـــة الأمثَلـــة الحســـابية (AOA) وخوارزميـــة الغَمْـــر العظـــيم (GDA)، فــي محاولـــةٍ لمواجهـــة تحـــدّي التّشـــخيص. وقـــد تـــمّ اختبـــار نمـــوذجيْن: (AOA) و (AOA-GD) علـــى 23 مـــن مجموعـــات البيانـــات المرجعيـــة الطبّيـــة. وبنـــاءً علــى جميـــع البيانـــات التّجريبيـــة، فـــإنّ تهجـــين الأمثَلـــة الحســـابية بخوارزميـــة الغَمْـــرالعظيم زاد علــى نحـــوٍ ملمـــوس مـــن قُـــدرة الأمثَلـــة الحســـابية علــى البحـــث. بعدئـــذٍ، تمّــت مقارنـــة النّظــام الهجـــين المقتـــرح بـــاثنتيْن مــن طُــرق انتقــاء السِّمات المســـتخدَمة ســـابقاً (CHIO-GC) و (LBMFO_V3). وعنـــد تطبيـــق الطّريقـــة المقترحـــة فــي هـــذا البحـــث والطّــريقتين السّـــابقتين المشـــار إليهمـــا آنفـــاً علــى مجموعـــات البيانـــات الـــثلاث والعشـــرين، حقّقـــت الطّريقـــة المقترحـــة معـــدّل دقّـــةٍ بلـــغ 0.80، وتجـــاوزت بذلك الطّريقتين السابقتين.

CERVICAL CANCER DETECTION AND CLASSIFICATION USING MRIs

Ichrak Khoulqi and Najlae Idrissi

ABSTRACT

Cervical Cancer (CC) is the second most frequent malignancy in women worldwide, with a 60 % mortality rate; it is the leading cause of death worldwide. The majority of cervical cancer deaths occur in less developed countries where there is a lack of screening programs and sensitization about the disease. CC cannot be detected in its early stages, since it does not reveal any symptoms and has a long latent period. Accurate staging can aid radiologists in providing effective therapy by utilizing diagnostic methods such as MRIs. In this paper, two approaches are proposed. The first consists of introducing an automatic system for early detection of CC using image processing techniques and axial, sagittal T2-weighted MRIs for analysis to determine the pathological stage of tumour and identify the real impact of cancer, that will help the patient to be treated with high efficiency and properly. This detection process goes through three major steps; i.e., pre-processing to make the representation of MRIs significant and easy to be analyzed, then segmentation was performed by region growing and geometric deformable techniques to extract the region of interests (ROIs). In the next step, we extract two categories of features based on statistical and transform methods in order to describe our ROIs. At the final step, five classifiers were trained to classify the MRIs into two classes: benign or malign. The second approach aims to increase the performance of pre-trained Deep Convolutional Neural Networks (DCNNs) based on Transfer Learning (TL) used to classify our Female Pelvis Dataset (FP_Dataset) by adopting the stacking generalized method that provides a more efficient and robust classifier. Data augmentation is a pre-processing method applied to our MRIs and a dropout layer is used to prevent networks from overfitting in our small dataset. The results of experiments show that data augmentation and stacking generalization represent an efficient way to improve accuracy rate of classification.

KEYWORDS

1. INTRODUCTION

Cervical cancer is ranked 4th among the world's female cancers. About 500 000 cases of cervical cancer [1], 200 000 cases of cancers of the body of the uterus and 200 000 cases of other gynecological cancers, such as ovaries, vulva and vagina, appear each year [1]. In Morocco, uterine cancer is ranked 2nd among female cancers in Moroccan women, with more than 3300 new cases and 2500 deaths per year [2].Cervical cancer is characterized by uncontrolled proliferation of abnormal cells that can invade and damage normal tissue [3]. The majority of cervical cancers originate in the cells that line the cervix. These cells do not transform directly into cancer; instead, normal cells in the uterus progressively develop precancerous changes that can turn into cervical cancer. The incidence of cervical cancer increases with age and reaches a stage from age 50; the main cause of CC is due to a sexually transmitted infection: Human Papillomavirus (HPV).In most cases, this infection is eliminated naturally in about 80% of women and in 10% of women this virus can cause precancerous lesions that can develop into tumours. Other causes that can lead to cancer include smoking, suppression of the body's immune system…etc. As mentioned above, CC has no symptoms. Therefore, as part of prevention, women should perform a Pap Smear test, which is a widely used colposcopy to check the uterine and vagina. This test can identify abnormal and irregular cells on the cervix and helps detect the cancerous tissues at an early stage.

In this work, we will take the following technical route: we will represent the state-of-the-art in the related work section, the proposed approaches, which are divided into two subsections; the first depicts approaches based on image processing techniques and the second on Deep Convolutional Neural Networks (DCNNs), the evaluation of the different proposed approaches in the experimental results section and finally the conclusion.

I. Khoulqi and N. Idrissi are with DICC Team, Data4Earth Laboratory, Sultan Moulay Slimane University, Beni Mellal, Morocco. Emails: ichrak.khoulqi@gmail.com and idrissi.najlae@gmail.com

2. RELATED WORK

MR image interpretation for the diagnosis of female pelvis part is a time-consuming, sensitive and hard complicated task due to different characteristics, such as size, shape and texture. To overcome this problem, two different categories of pre-treating, segmenting and classifying techniques exist.

2.1 Classical Methods

To detect and classify Cervical Cancer (CC), several works have been carried out using colposcopy images to observe cell change and X-ray techniques to analyze tissues and tumour extension in case of their presence. Among these works, we find the following: Bethanney et al. [4] suggested an automatic cervical cancer diagnosis method based on texture descriptors and SVM multiclass classification (Support Vector Machine). Their work is divided into four stages: Pre-treatment, in which MRIs are pre-treated by removing undesired noises and other factors by equalizing the histogram and increasing the contrast, as well as using a nonlocal means filter. In the segmentation phase, the authors used the region growth method to extract ROIs; in the description phase, a gray level co-occurrence matrix was used to extract significant features about ROI texture; and finally, multiclass SVM was used to classify data into non-cancerous, benign and malignant.

Mithlesh et al. [5] work revolves around the use of Pap Smear images to perform automatic approaches to determine the shape, size and texture of the nucleus of the cell; their work is divided into three stages: Pre-processing, segmentation and classification. The Pap Smear images are converted to grayscale and in order to extract the nucleus from the cytoplasm, the borders must be highlighted with contrast enhancement before using the Gaussian filter to reduce noise. To extract the ROI, several thresholds are utilized that vary depending on the input image; the image gradient is calculated to define the boundaries of the nucleus and then the morphological operators are used to clean the segmented image. They extracted the nucleus properties for the classification phase during the description step and the characteristics used are: number of nuclei in the image, area, compactness, major axis, minor axis, aspect ratio and eccentricity. They used the Support Vector Machine to classify their data into three different classes, CIN1, CIN2 and CIN3.

Robert et al. [6] also employed Pap Smear images to construct an automatic detection system in order to eliminate difficulties caused by interpretation under the microscope; their work was divided into four phases: They used a Kernel of 3*3 for pre-processing and to determine the optimal kernel coefficients, they used a genetic algorithm with a repeat of 50 to 100 iterations, followed by a convolution procedure to obtain a noiseless image. After transferring the improved image (RGB) to the HSI space, the segmentation stage employs mathematical morphological operators. In the feature extraction or description step, five features are extracted: energy, local variation, correlation, entropy and homogeneity. Finally, in classification, they used the K-means method to classify the input image as normal or abnormal based on the extracted characteristics.

The work [7] comprises the identification and categorization of cervical cancer utilizing MRI scans, with the following workflow: Preprocessing; this phase allows for the enhancement of MRI image intensities *via* gamma correction and the probability of pixel luminance distribution. Then, the authors adopted Otsu thresholding for segmentation, which was based on thresholding by determining an ideal grey level to separate the ROI from its background. In the feature extraction step, they used gray level co-occurrence, contourlet features and Gabor features matrix. At least, the classification step consists in classifying the data into normal or abnormal based on the Support Vector Machine (SVM) technique. Sajeena et al. [8] proposed an automated method for the identification of cervical cancer by segmenting and classifying cervical cells; their paper is provided in four steps: Image acquisition for Pap Smears, then noise reduction using the non-local means filter to improve the visual quality of the image. Then, the K-means clustering method is used to partition the cell to cytoplasm, nucleus and background and the Radiating Gradient Flow (RGVF) snake is conducted. Six features are calculated from the extracted ROI: Area, compactness, major axis, minor axis, aspect ratio and nucleus homogeneity. Finally, classification is used to classify the data as normal or abnormal, with three different classifiers.

2.2 Deep-learning Methods

The last years, deep-learning model (Convolutional Neural Network) achieved a big success in

143

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 08, No. 02, June 2022.

analyzing and classifying biomedical images that were being seen in health-care systems for detecting tumours, like lung nodule classification, breast cancer detection, identification of skin disease, …etc. In order to get a stable, fast and accurate model Transfer learning technique is used to improve CNN architecture.

In [12], Almubarak et al. used CNN to classify squamous epithelium into 4 grades of Cervical Intraepithelial Neoplasia (CIN) by dividing the epithelium into 10 segments and each segment into 3 parts (top, middle, bottom); then, they used a Deep Neural Network which consists of 32 filters with limited epochs and obtained a gained accuracy rate of 77,25%. Sun et al. [13] adopted random and hybrid decision tree to classify the cancerous images (Pap Smear images) into two classes: normal or abnormal, where the accuracy rate was 94.4%.

Antonios Makris et al. [14] proposed using Deep Convolutional Neural Networks to classify COVID-19 using a variety of pre-trained models including Xception, InceptionV3, MobileNetV2, VGG16, VGG19 and NasNetLarge. The dataset is divided into three classes: Covid, Normal and Pneumonia, which are classified using chest X-ray images that were split into 80 % and 20% for training and testing, respectively, with 35 epochs, a learning rate of 1e-3 and a batch size of 8.The best accuracy was 95.88% when using VGG16 as a fine-tuned pre-trained model.

The authors of [15] used deep learning to create a diagnostic system for cervical squamous intraepithelial lesions. They collected data to create three deep-learning models. The first step in this work consists of pre-processing data by resizing all images into 512*512 pixels and dividing them into 100 categories by K-means and randomly relocating them into three sets (training set, validation set and test set). The second step is based on transfer learning technique using ResNet pre-trained model on ImageNet dataset to improve the efficiency of the proposed techniques, in order to classify images into two groups. Their proposed approach obtained an accuracy rate of 84.10% with an AUC of 0.93 and for evaluating segmentation, the DICE metric was used and they obtained an average accuracy of 95.59%. The authors proposed two deep-learning CNN architectures to detect cervical cancer based on colposcopy images in [15]. The first is the pre-trained VGG19 model, which achieved an accuracy of 73.3% and the second is CYENET, which achieved an accuracy of 92.3%.

In [9], Saumaya et al. improved and fine-tuned the EfficientNetB3 model to classify malignant skin lesions by comparing several pre-trained models, such as ResNet50, InceptionV3, InceptionResNetV2 and EfficientB0-B2 models. Their proposed model achieved an accuracy of 87.12%, a recall of 87.12%, a precision of 87.12% and an F1-score of 85.12%. In [10], a modified K-means is suggested to extract the region of interest from mammography, followed by textural features retrieved using GLCM and Gabor and CNN features extracted using the pre-trained model InceptionV3 after simple pre-processing and cropping. Cross-validation is used to assess the quality of the retrieved features using five different classifiers; namely, SVM, KNN, MLP, RF and NB.

3. THE PROPOSED APPROACHES

The proposed approaches are presented in this section by explaining each stage of the computer-aided diagnostic system of the first approach that has some objective to finally achieve the obtained results and those can be obtained by adopting and applying different techniques, such as noise removal, filtering and contrast enhancement in the pre-processing step. Similarly, the image segmentation step consists of extracting the ROIs, followed by feature extraction and selection to obtain significant characteristics that better define our ROIs. Finally, we go to the classification step to categorize our female pelvis dataset as malign or benign. Due to a lack of female pelvis MRIs, we chose a deep convolutional neural network with transfer learning as our second strategy in order to increase the learning capability of the suggested model and develop a meta-learner that improves the performance of our proposed model. In the next subsections, we will go through each stage of each proposed approach in further detail.

3.1 First Proposed Approach

To increase the chances and possibilities of detecting cervical cancer in early stages, we have proposed two approaches based on the analysis and interpretation of MRI images. The female pelvis MRIs used in this work are collected from different web sources. In this paper, we worked with two slices of the

pelvic part: axial and sagittal, which allowed us to better visualize the tumour in order to extract it and place it in its most appropriate pathological stage, with the aim of enhancing the chances of survival and healing.
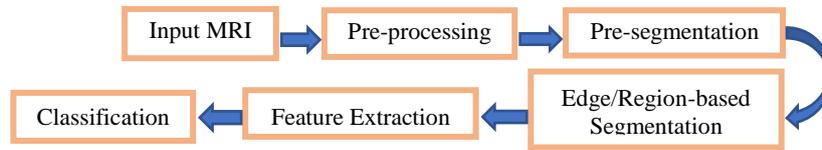

Figure 1. Block diagram of the first proposed system.

Figure 1 depicts the global schema of our first proposed detection system based on image processing techniques taking place in four steps.

1) **Pre-processing:** Denoising and enhancing borders of structures and tissues in order to get a better visual quality of MRI scans for further phases.
2) **Segmentation:** Consists of extracting the region of interests (ROIs) or tumoral zones using two different approaches: Distance Regularized Level Set Evolution (DRLSE) and Region Growing-based Gradient. Those techniques are largely used for image analysis in many fields, such as segmentation of brain tumours using LevelSet evolution [16], automatic detection of man-made objects from aerial and satellite images based on LevelSet evolution [17], automatic image segmentation by integrating color-edge extraction and seeded region growing [18] and adaptive region growing technique using polynomial functions for image approximation [19].
3) **Feature Extraction:** Crucial step for separating beneficial characteristics from the extracted ROIs.
4) **Classification:** Final step of the process of detection and classification by categorizing our female pelvis dataset into malignant or benign.

Detection of cervical cancer is a complicated task that requires high precision in detection to stadify our Female Pelvis MRIs, which is why we have proceeded with a pre-processing step to improve the MRIs quality and then we select our ROIs as is required in the segmentation step using two different approaches as cited above; region-based technique by implementing region growing method, which is simple, fast and computationally inexpensive. The second segmentation strategy was based on boundary or contour to recognize abrupt changes in grayscale images. These algorithms are versatile in recognizing ROIs by respecting their morphology that will help us in further steps.

3.1.1 Pre-processing

MRI images of the pelvic area are altered by a variety of noise leading to inaccuracy in the detection of cervical cancer and this is due to a variety of factors during the process of imaging the pelvis, such as the influence of contrast, temperature and other factors like the technics problem in the machines during the capturing which affects the MRI by a blurring which does not allow a better visualization of the different regions of the pelvis. In order to remedy these alterations and enhance the visual quality of images to be more useful and usable and to increase the efficiency of our different proposed approaches, we have to go through an essential phase which is pre-processing. In this study, it is essential to have information on the edges to be able to retrieve the region of interest (ROI); for example, we have filters such as the averaging filter that smoothes the sudden change in illuminance on the borders, which is not perfect for our case study, so we chose to use a non-linear filter that removes Gaussian noise while keeping the edges, which will give us a clearer image that is easier to analyze and interpret. It is the bilateral filter.
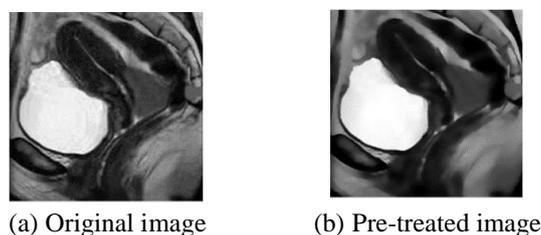

(a) Original image   (b) Pre-treated image

Figure 2. Pre-processing phase based on bilateral filter.

Bilateral Filter (BLF) [21]-[22]: The basic idea of this filter is to add weights to the Gaussian convolution taking into consideration the distance between pixels in the illuminance space. This filter is used for noise reduction and image enhancement; this filtering technique is adapted to our MRI images by preserving their disjunctions to separate the images into different regions [16]. It consists in filtering the MRI images of the pelvis while keeping the abrupt change of intensities between the MRI structures (the edges).

Contrast Adjustment

MRI images of the pelvic area are pre-treated with the bilateral filter (BLF) to remove the noise existing in the images while keeping the edges of the different parts composing the pelvis. Then, we move on to another problem that arises when taking MRI scans, which is the contrast. To limit the impact of this problem, we must adjust it to improve the clarity of our data and facilitate the execution of the next steps of our proposed detection system. MRI images are enhanced using an innovative strategy of adjusting the image intensity; the imadjust function is used to transform the intensity of images to enhance the appearance of the pelvis regions for better quantification results. This function is given by $J = imadjust(L, [Low_{in} \ High_{in}], [Low_{out} \ High_{out}], gamma)$ in Matlab [23]; it consists into mapping the intensity values for the input image to new values (J between $Low_{in} \ and \ High_{in}$ mapped to values between $Low_{out} \ and \ High_{out}$); as a default, 1% of the data is immersed at both low and high input image intensities [24]. If gamma is greater than 1, the mapping goes toward dark output values. In this work, gamma value is set to 1 which is the default value and it gives desired results for our MRI data of the pelvis.

Pre-segmentation: Initialization Phase

In the field of cervical cancer detection, the authors proposed in this work [25] an initialization phase before proceeding to the segmentation based on the K-means algorithm. Using this algorithm, we observe that the different structures of the pelvis are better contrasted and the borders of the neighboring structures are clear. For the axial oblique section, it is divided into 5 essential regions (uterus, rectum, bladder, pelvic floor muscles and the tumour if present) as well as for the sagittal section (uterus, ovaries, rectum, acetabulum and the lesion if present).Figure 3 illustrates the results obtained for the pre-segmentation phase.

Segmentation Step

Image segmentation is a critical and crucial step in image analysis, especially for medical images where the information to be extracted is very important and where any loss of information can modify the final decision. It consists of dividing an image into regions or categories, which correspond to different objects or parts of objects. The interest of this segmentation is to be able to manipulate these regions *via* high-level processing [26] to extract their shape characteristics (i.e., distance, position, size, …etc.). Segmentation methods are classified according to two properties: similarity and discontinuity. Based on these properties, image segmentation is defined by two categories: edge-based segmentation and region-based segmentation. The region-based segmentation divides an image into similar areas of connected pixels. In our study, we were based on the region growing method and for edge-based or discontinuity-based techniques, we were based on the regularized distance of level set evolution. In the segmentation step of our proposed detection system, we extract the ROI which is the cervical tumoral zone, the region where the cancer is present and where it has spread too. More details are given in the following sections.
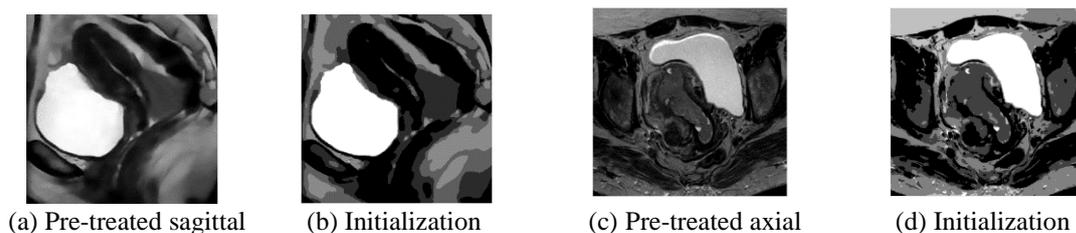

(a) Pre-treated sagittal   (b) Initialization   (c) Pre-treated axial   (d) Initialization

Figure 3. Sagittal and axial image initialization by K-means algorithm.

Geometric Deformable Model

Geometrical method is a model based on the curve evolution technique. The curves are always

evolving towards the normal direction. Geometrical snakes are represented implicitly [24] as the zero-level set of the surfaces with higher dimensions [27]. The update is performed on the surface function in the entire image domain.

Level Set Method (LVS): LVS is presented by Osher and Sethian [28]. It is the surface that intersects the plan and gives us a contour. LVS is the formulation of active contours; this surface is updated with forces derived from the images. In 2D, the curve around the object to be segmented or extracted evolves depending on the internal and external forces.

- **Internal forces:** Defined in the properties of the curve; preservation of the smooth appearance of the curve during its deformation.
- **External forces:** Defined from the image; curve deformation according to the characteristics of the image.

In 3D, the level set (LSF) function incorporates this curve as a zero-level set, which means that it is a surface $\emptyset$ with no height $\emptyset = 0$. Moving fronts are noted by C represented by the zero level $C(t) = \{(x, y)/\emptyset(x, y, t) = 0\}$ of the level set function $\emptyset(x, y, t) = 0$. The main idea of this technique is to place a contour in an image which deforms until it achieves an optimal position and shape. A point $x = (x, y)$ belonging to the front of the surface evolves over time, so we denote by $x(t)$ its position. For each point $x(t)$ on the front has by definition no height, thus:

$$\emptyset(x(t), t) = 0 \tag{1}$$

Level set methods are represented by a partial differential equation (PDE) to determine the position of $\emptyset$ at any time t, where the PDE is as follows:

$$\frac{\partial \emptyset}{\partial t} + F|\nabla \emptyset_t| = 0 \tag{2}$$

The level set method is used to solve several problems in different research fields, such as medical imaging, engineering fields, …etc. The application of LevelSet Standard methods suffers from irregularities of the Level Set Function (LSF) during evolution. The PDE can develop a sharp or flat shape during evolution, which makes the calculations very inaccurate that necessitates a reinitialization; the latter is executed by periodically stopping the evolution and reshaping the degraded LSF as a signed distance function; the distance signed $z = \emptyset(x, y)$ is a surface the plane tangent of which makes an angle of 45° with xy in the plane and the z axis. This condition is verified by the property of the distance signed $|\nabla \emptyset| = 1$ [29]. The disadvantage of a reinitialization is that it affects numerical accuracy, which is why PDE is converted into a variational levelling [30] technique based on energy minimization, useful for adding external shape, color or texture information to the model.

Distance Regularized LevelSet Evolution (DRLSE)

For the segmentation of MRI images of the pelvic part, we use information provided by the edges to determine the external energy; the distance regularization method for the evolution of levelSet (DRLSE) [30] consists in determining a convolution function that smoothes and reduces the noise in the image. This function g always takes minimum values on the edges of an object against other regions. The formula of the convolution function is represented as follows:

$$g \triangleq \frac{1}{1 + \nabla G_\sigma * I} \tag{3}$$

where G_σ is the Gaussian kernel with a standard deviation σ and I is the image to be segmented defined on a domain Ω. The definition of functional energy is as follows:

$$\varepsilon(\emptyset) = \mu R_p(\emptyset) + \lambda L_g(\emptyset) + \alpha A_g(\emptyset) \tag{4}$$

where $\mu > 0$ is a constant and $R_p(\emptyset)$ is the term of regularization of level set; its formula is defined as follows:

$$R_p(\emptyset) \triangleq \int p(\nabla \emptyset) dx \tag{5}$$

where p is a potential; $\lambda > 0$ and $\alpha \in R$ are the coefficients of the functional energies $L_g(\emptyset)$ and $A_g(\emptyset)$.

The DRLSE (Distance Regularized Level Set Evolution) approach is used for the segmentation and extraction of the region of interest. In our case study, it shows a great efficiency, since it respects the edges and the geometric shape of the ROI given the high importance of the latter for staging. The

obtained results for segmenting axial and sagittal MRIs of the cervix by the DRLSE are shown in Figure 4 and Figure 5, respectively.
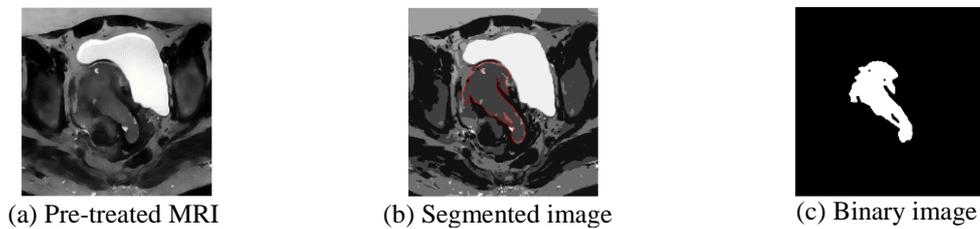

(a) Pre-treated MRI   (b) Segmented image   (c) Binary image

Figure 4. Segmentation of axial MRI by DRLSE approach.

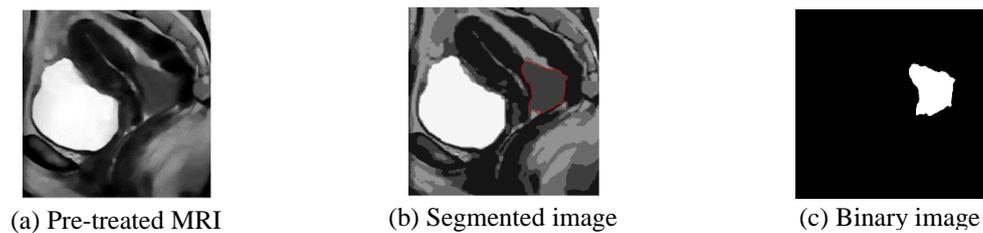(a) Pre-treated MRI   (b) Segmented image   (c) Binary image

Figure 5. Segmentation of sagittal MRI by DRLSE approach.

As shown from Figures 4 and 5, the pelvic parts or the suspected tumoral cervix zone is well limited by the red curve as presented in the figures, which delimits and fits too well the cervical tumoral parts to keep only the extracted ROI. As shown in Figures 4 (c) and 5 (c), we apply a simple thresholding to previously segmented images. We can notice from these results that the segmentation of cervical MR images by DRLSE approach gives a good separation of ROI with fine and better localized contours.

Region-based Techniques

Region growing method allows to group in an iterative way connected regions whose union respects a property of homogeneity; it is a tool used for image segmentation introduced by ZUCKER and it has been used repeatedly for the segmentation of medical images [17]. In this paper, new algorithm of segmentation based on the gradient and the seeded region growing for cervical cancer is introduced to extract the malignancy from the MRI images, so first we calculate the gradient of the pre-treated and initialized MRI in order to determine the edges of different parts that compose the MRI perfectly and to avoid the problem of over-segmentation caused by the selection of seed points. There are several edge detection algorithms; the most popular and used are Canny, Sobel and Prewitt. Commonly, the gradient of an image is computed by convolving the image with kernel (filter mask) yielding the image derivatives in x and y directions. The magnitude and direction of the gradient can then be calculated using those derivatives. The region growing technique is our second approach, since it allows us to segment the image well and extract the lesion with precision, always keeping its geometric shape as we have already mentioned in the first approach, since the geometric shape is very important for the next phase (classification). The results of this approach are illustrated in Figures 6 and 7 for axial and sagittal MRI cervix images, respectively.
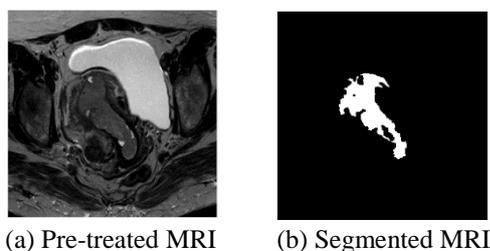

(a) Pre-treated MRI   (b) Segmented MRI   (a) Pre-treated MRI   (b) Segmented MRI

Figure 6. Segmentation of axial MRI by   Figure 7. Segmentation of sagittal MRI by
      RG-based gradient.   RG-based gradient.

3.1.2 Feature Extraction

In order to extract beneficial characteristics from the image, we should go through feature extraction

or descriptors, where selecting features helps in the classification, clustering or prediction step by representing our data in a better way .It consists in finding the most compact and informative set of features by measuring properties like color, texture or shape of the whole image or sub-image that is represented in our study by the region of interest (ROI) or the tumoral zone.

During this work, we choose Gray Level Co-occurrence Matrix (GLCM) [31], which is a **statistical tool** that extracts second order texture data from image as texture analysis is well adopted to the monitoring of disease and for characterizing the lesions [32].Also, we proceed to the utilization of Local Binary Pattern (LBP) that characterizes the texture present in the image in the gray levels by attributing to each pixel of the image a value that describes or characterizes the local binary pattern around this pixel [33].Among the approaches of texture analysis, there are techniques based on **transform methods**, such as Discrete Cosine Transform (DCT) [34] that allows the change of the field of study while keeping exactly the same studied function. In our work, we study MRIs; which means a 3-dimensional function: X and Y, indicating the pixel and Z the value of the pixel at this point. The features extracted from the region of interest (ROI) are depicted in the following table.

Table 1. Summary of image features extracted.

Feature	Feature description
DCT	a1, a2
LBP	l1, l2, l3, l4, l5, l6, l7, l8, l9, l10
GLCM	g1, g2, g3, g4, g5, g6, g7, g8
MAX	Length of major axis of the extracted ROI
A	Total number of pixels in the ROI
C	Compactness$(\frac{P^2}{A})$, where P is the perimeter and A is the area of the ROI

3.1.3 Classification Step

Classification plays an important role in medical applications. In this section, we have outlined the classification of cervical cancer. The CC detection is a very challenging task as this cancer occurs without revealing any symptoms. In this paper, five classifiers including **Random Forest (RF)** [35], **Artificial Neural Network (ANN)** [36], **Decision Tree (DT)** [37], **K-nearest neighbour (KNN)** [38] and **Gradient boosted tree (GBT)** [39] are applied to figure out the appropriate classification of our data. Before the step of classification, we should go through the normalization, because it makes data have the same scale, so each extracted feature is equally important; if we forget to normalize, one of our features might completely dominate others. In our case, we choose using **Z-score normalization** by adopting the following formula:
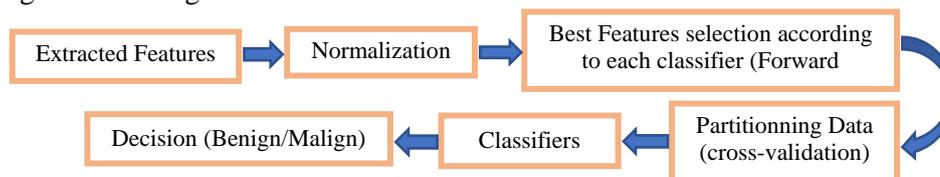
Figure 8. Block diagram of the proposed approach of classification.

$$Z_{score} = \frac{value - \mu}{\sigma}$$

(6)

where μ is the mean value of the feature and σ is the standard deviation of the feature. This technique shows a good result, as it can handle outliers. The schema below depicts the CC classification process.

3.2 Approaches to Classification-based Deep Learning

The second proposed approach is based on Convolutional Neural Networks (CNNs). It consists of adopting an ensemble model [46] designed by collaborating six different transfer learning techniques; i.e., NasNetLarge, Xception, VGG16, ResNet152V2, MobileNetV2 and InceptionResNetV2. Ensemble learning leads to augment the classification rate, which leads us to get a powerful approach by remedying the problem caused by the high variance that occurs from the stochastic nature of the neural networks in training. The ensemble learning approach proposed in this work is called stacking generalization or stacking, where a new model learns how to better combine the predictions from the different sub-models. It enhances the learning system's generalization ability; the base learners or

149

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 08, No. 02, June 2022.

classifiers can be chosen or generated in two manners: The dataset is identical to the various learning algorithms (heterogeneous learners) or the other way around (homogenous learners). The stacking procedure goes through two steps:

- **Level 0:** learns to make predictions from training dataset (inputs).
- **Level 1:** takes the outputs of the proposed sub-models as input and the meta-learner makes predictions from this data.

3.2.1 Convolutional Neural Networks

CNN models perform in a better way in various domains, such as industry, agriculture, detection and classification of medical diseases. The conception or the architecture of CNNs clones the visual cortex system of humans [48]-[49]. It consists of three main steps; the first is convolutional layers, the second is the pooling layer and finally the fully connected layer that are mandatory layers; the secondary layers are depicted by the dropout layer which is used to overcome overfitting during training and the normalization layer; those layers are exploited to increase CNN model performance. The two first layers (convolutional and pooling layers) are in charge of learning to extract the essential features by applying a window size named kernel that convolves the image to extract significant and important features that will be used in the classification tasks and the last one (fully connected layer=FC) does the classification.

In this paper, we took six different types of Deep Convolutional Neural Networks (DCNNs) as cited above that have been pre-trained on natural image dataset (ImageNet) whose weight is used in Transfer Learning experiments. The Softmax function is implemented at the final layer (FC) to output the predicted probabilities to determine the class of the female pelvis MRIs.

3.2.2 Transfer Learning

It is often difficult to get a large dataset in the medical Imaging field; the data for CC_MRIs is much lesser. DCNNs cannot learn in a better way the small Dataset; it requires a huge one to train the model in order to provide better results. DCNNs with small dataset lead to overfitting; it occurs when the noise of data is captured by the model and it arises when this latter fit the data too well due to small dataset. All these problems can be resolved by Transfer Learning (TL) techniques.TL uses pre-trained DCNNs in two ways; first, they are used for feature extraction and then the important knowledge is guarded and used by another model designed for classification. The second one is a more sophisticated technique, where some specific modifications are made to the pre-trained model to achieve good results. These modifications include architecture and hyper parameter adjustment. It is important to use the knowledge gained when we solve a problem of natural image recognition to solve different problems of medical image classification. TL or fine tuning is used to enhance the efficiency of DCNNs as cited above; it comprises removing the last fully connected layer of the pre-trained model on ImageNet, since the outputs are 1000 classes to adapt them to our desired output (two classes= malign/benign).

In this proposed approach, we use Adamx optimizer and the categorical cross-entropy loss function to train the model on our MRIs.We adopt this TL technique, because our female pelvis dataset is small and this latter will help us in improving classification rate to get an efficient model of female pelvic MRI classification. To treat the lack of screening MRIs, we use an important technique which aids in enhancing our dataset; it is data augmentation.

3.2.3 Data Augmentation

Data augmentation is the most common method that reduces overfitting. It is used to artificially expand the size of dataset by creating modified versions of MRIs to improve the ability of the fit models to generalize what they have learned from new images. In this paper, the approach of augmentation is a pre-processing step, which means a geometrical transform of the image, such as random flip horizontally and vertically, random rotation within 45°, horizontal shear within 0.2 times, the image width, height and zoom within 0.2 times.

3.2.4 Methodology of the Second Proposed Approach

The scheme in (Figure 9), illustrated a diagram of the second proposed pipeline by presenting how we stacked different machine learning algorithms to improve the prediction rate for FP_Dataset.
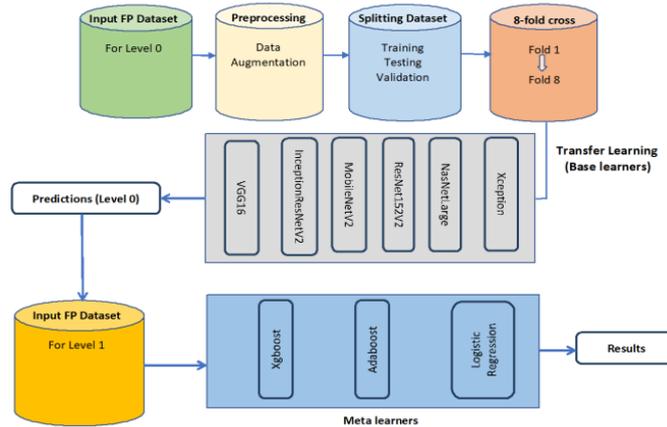
Figure 9. The second proposed pipeline's detailed system.

The following algorithm summarizes the different steps of our proposed approach for FP_Dataset classification.

Algorithm1: Pseudo-code of the Stacking Generalization based Cervical Cancer Classification
1: **Input:** Training FP_Dataset with Malignant and Begnin cases D= { $(x_1, c_1),(x_1, c_1), …. ,(x_n, c_n)$ }
2: Base level classifiers $C_1, … , C_k$
3: Meta level classifiers $\hat{C}_1, \hat{C}_2, \hat{C}_3$
4: **Output:** Trained ensemble classifiers $\widehat{M}_1, \widehat{M}_2, \widehat{M}_3$
5: **BEGIN**
6: **Step1:** Train base learners by applying classifiers C_i to FP_Dataset.
7: **for i**=1, …, k **do**
8: $B_i = C_i(D)$
9: **end for**
10: **Step 2:** Construct new Dataset of predictions $\widehat{D}$
11: **for j=1, …., n do**
12: **for i=1, …, k do**
13: **% use** B_i to classify training sample x_j
14: $z_{ij} = B_i(x_j)$
15: **end for**
16: $\widehat{D}$= $\{Z_j, c_j\}$, where $Z_j= \{z_{1j}, z_{2j}, …. , z_{nj}\}$
17: **end for**
18: **Step 3:** Train a Meta level classifiers
19: $\widehat{M}_1 = \hat{C}_1(\widehat{D})$
20: $\widehat{M}_2 = \hat{C}_2(\widehat{D})$
21: $\widehat{M}_3 = \hat{C}_3(\widehat{D})$
22: **Return** $\widehat{M}_1, \widehat{M}_2, \widehat{M}_3$

In the following paragraphs, we provide a summary of the different deep convolutional neural Networks used in this paper.

Xception: The Xception network was introduced by Chollet in 2017 [47]. It is a DCNN that stretches the inception concept to extremes; it introduces new inception layers that are created by depth-wise convolutions that are alternatives to classical convolutions and much more efficient in terms of computation times. The input format of images is 299*299. This network has a depth of 126 with 36 convolutional layers to extract features and a global average pooling layer is adopted to replace the fully connected layers in order to reduce the number of parameters. Then, the activation function Softmax is used to output predictions.

VGG16: VGG16 is a DCNN proposed by K. Simonyan and A. Zisserman (the visual geometry group). It has a depth of 16 [40] weight layers; it achieves an excellent accuracy on the ImageNet classification and on image recognition dataset. The input images format is 224*224, then the filters of 3*3 are applicated in all convolutional layers in order to reduce the number of parameters.

MobileNetV2: MobileNetV2 is an architecture proposed by Google to run on mobiles and an

151

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 08, No. 02, June 2022.

embedded system [11]. In order to reduce the trainable parameter number, depth-wise separable convolutions are adopted in MobileNet architecture. Depth-wise separable convolution divides the kernel into two small kernels; one for depth-wise convolution and the other for point-wise convolution. This technique of dividing kernels helps in reducing the computational cost. In this work, the pre-trained MobileNet model on ImageNet dataset is imported. It uses an input image of 224*224 and a depth of 88.

InceptionResNetV2: InceptionResNetV2 is an architecture based on InceptionV3 [12] and Microsoft's ResNet [13]-[14]. It combines the properties of both. This architecture is used as a high-level feature extractor which provides information about the image content that can help to characterize the information contained in our MRIs to better classify them. This model is adopted for extracting features and for classification. It has an input image format of 299*299 and a depth of 572.

NasNetLarge: NASNet architectures can be tuned using the reinforcement learning search method by introducing a new concept of normal cell and reduction cell. The NasNetLarge architecture is designed specifically to train on large datasets. Since training on large datasets is expensive and resource-intensive, the search for an architectural block is performed on a small dataset and then the block is transferred to a larger dataset using the NASNet search space. The key aspect of NasNetLarge includes the **ScheduledDropPath** regularization technique that significantly improves the generalization of NASNet models.

ResNet152V2: Residual Networks (ResNet) [50] were proposed as a family of multiple deep neural networks with similar structures, but with different depths. ResNet introduces a structure called residual learning unit to alleviate the degradation of deep neural networks. The main merit of this unit is that it produces better classification accuracy without increasing the complexity of the model. The difference between ResNet152V2 and the original V1 is that V2 uses batch normalization before each weight layer. In the field of image recognition, it has a strong performance that justifies our selection among ResNet by Resnet152V2 as it achieves the best accuracy.

We have modified the pre-trained architectures of Xception, VGG16, MobileNetV2, InceptionResNetV2, NasNetLarge and ResNet152V2 by adding dense layers with 'relu' activation, dropout and SoftMax layers with two outputs (malign/begnin). We implement our methods using Keras library and Tensorflow. For our experiments, we load weights of the pre-trained CNNs given by Keras. In the following part, we focus on machine-learning algorithms adopted as Meta learners to increase the classification rate.

Adaptative Boosting: Adaptative boosting (Adaboost for short) algorithm is a technique used as an ensemble method in machine learning. The reason behind the name of this algorithm is that the weights are re-assigned to each instance by attributing higher weights to incorrectly classified instances. It is used to reduce bias as well as variance for supervised learning. As we said above, the main idea is to give more importance to the misclassified data points built by the first weak learner and then construct another weak learner based on the incorrectly classified data. The following formula represents the Adaboost algorithm:

$$F(x) = sign(\sum_{k=1}^{K} \theta_k f_k(x)) \tag{7}$$

where, K: weak classifiers' number, θ: Weight of k-th weak classifier and $f_k(x)$: weak classifier.

Logistic Regression: Logistic regression is a linear model with binary output; it models the variable with a line for two dependent variables or hyperplane in case of more than two variables. The logistic operation is as follows:

$$p = \frac{1}{1+e^{-(\beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n)}} \tag{8}$$

where, p: the probability of success (the presence of malignancy), β_0: the model intercepts and β_i: the regression coefficients.

eXtreme Gradient Boosting: eXtreme Gradient Boosting (XGB for short) is one of the approaches to construct ensemble-learning models. It is a tree ensemble model that is represented by a set of classification and regression trees (CARTs). The mathematical formula can be written in the following form:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i) \tag{9}$$

where, K: the trees number, f: Function in functional space F and F: Set of all possible CARTs.

The objective function that needs to be optimized (minimized) can be represented as follows:

$$obj(\theta) = \sum_i^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \tag{10}$$

where, $l(y_i, \hat{y}_i)$: training loss function and $\Omega(f_k)$: regularized term.

4. EXPERIMENTAL RESULTS

4.1 Evaluation of Classical Methods

To evaluate the performance of our computer-aided diagnostic system for cervical cancer (CC) detection and classification, we proceed by evaluating the segmentation and classification proposed techniques.

4.1.1 Quantitative and Qualitative Evaluation of Segmentation

The segmentation process is critical and must be completed appropriately by delimiting the target ROI. To validate this, we begin with a visual inspection, as illustrated in Figure 10. Respectively, the figure represents the segmentation of axial and sagittal MRI cervix images. The first column (a) represents the original cervix images to be segmented, while columns (b) and (c) represent the segmented images obtained by using KM-DRLSE and KM-RG, respectively.

As reported previously, the two approaches give better segmentation by respecting and keeping well the geometric shape of ROI, although the region growing approach outperforms the DRLSE method by specifying the fine details of ROI structure as well as the affected tissues. The second evaluation relies on numerical evaluation based on SSIM (the structural similarity measure [42]) and ZSI (the Zijdenbos similarity index [43]) parameters.

The measurements obtained for these two indices are given in the following table.

Table 2. Evaluation of segmentation approaches.

Experiments	DRLSE		RG	
	SSIM	ZSI	SSIM	ZSI
Experiment 1	0,9429	0,8840	0,9306	0,8067
Experiment 2	0,9863	0,7629	0,9845	0,5756
Experiment 3	0,9535	0,8948	0,9524	0,8726
Experiment 4	0,9359	0,8434	0,9269	0,7846
Experiment 5	0,9700	0,8889	0,8684	0,8684
Experiment 6	0,9664	0,7479	0,9650	0,7386

(a) Original MRI   (b) Segmented MRI by KM-DRLSE   (c) Segmented MRI by KM-RG
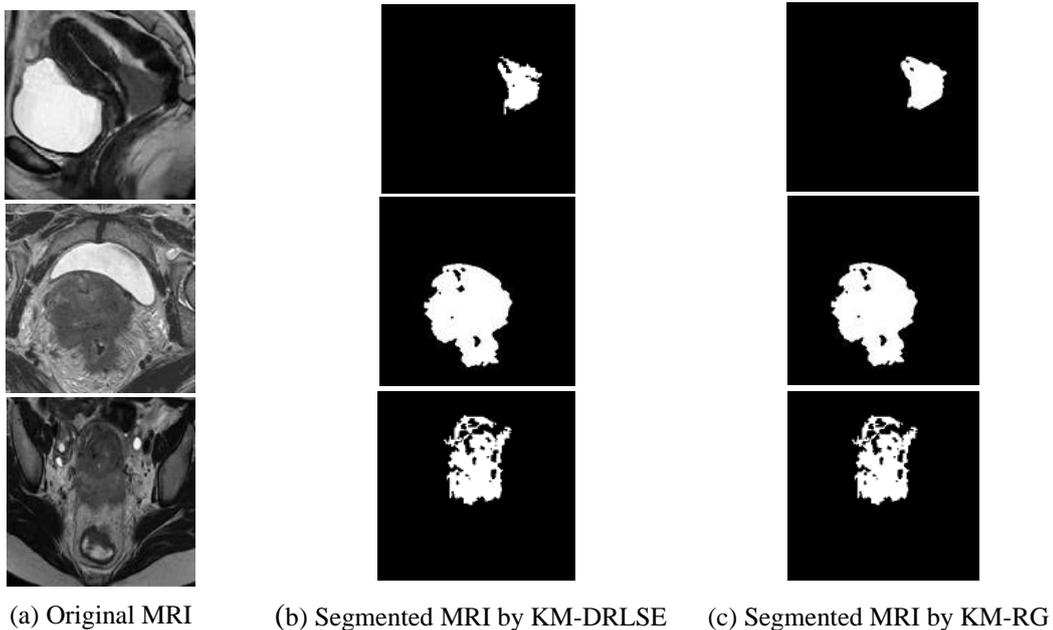
Figure 10. Segmentation results by the proposed approaches.

153

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 08, No. 02, June 2022.

SSIM and ZSI are close to one, indicating better segmentation accuracy. Table 2 shows the results obtained by using the cited evaluation metrics based on the proposed approaches KM-DRLSE and KM-RG. All of the similarity measures produce significant results for all of the values, but as shown in the table, KM-DRLSE outperforms KM-RG, indicating that KM-DRLSE is a powerful technique for extracting a tumoral zone from pelvic MRIs.

4.1.2 Evaluation of Classification

In this sub-section, we will evaluate the classification of pelvic MRIs. After selecting the best feature for each classifier, we proceed to the step of determining the measure ratings in order to assess classification. The results obtained for classification based on KM-DRLSE and KM-RG, respectively, are represented in Tables 3 and 4.

Table 3. Evaluation of classification using KM-DRLSE.

Classifier Performance	RF	GBT	KNN	DT	ANN
Accuracy	93,103%	93,103%	89,66%	86,21%	86,21%
Recall	90%	80%	90%	80%	90%
Specificity	94,74%	100%	89,47%	89,47%	84,21%
Precision	90%	100%	81,82%	80%	75%
F-measure	90%	88,89%	85,71%	80%	81,82%

Table 4. Evaluation of classification using KM-RG.

Classifier Performance	RF	GBT	KNN	DT	ANN
Accuracy	79,31%	79,31%	86,21%	86,21%	86,21%
Recall	60%	60%	60%	80%	80%
Specificity	89,47%	89,47%	100%	89,47%	89,47%
Precision	75%	75%	100%	80%	80%
F-measure	66,67%	66,67%	75%	80%	80%

The RF and GBT classifiers perform the best in the approach based on KM-DRLSE with 93,103% of accuracy, while the approach based on KM-RG gives us a good result by using ANN, DT and KNN classifiers.

4.2 Evaluation of the Second Proposed Approach

In this sub-section, we evaluate the second proposed approach adopted to classify FP_Dataset by representing our experimental results.

Cross-validation is a method of evaluation consisting in preventing overfitting and improving the model performance during evaluation. We evaluated the classification performance of the fine-tuned DCNNs with eight-fold cross-validation. This technique allows testing all of the datasets each with a total of eight verifications, which means that the 8th group is used for validation and the other seven groups combined are used for training. To provide comparable results, the same eight training and testing sets were kept in each method.

4.2.1 Evaluation Metrics

To quantify the model's classification performance, evaluation metrics are used. Confusion matrix (CM) was employed, which included accuracy, precision, recall, specificity and F1-score. We calculate the following quantities while evaluating these measures: True Positive (TP), False Negative (FN), True Negative (TN) and False Positive (FP).

After loading the image into the Python Image Library (PIL) format, we convert the PIL image into a Numpy array and we prepare inputs of shape (3,224,224), (3,299,299) and (3,331,331) according to each pre-trained CNN using Keras tools of pre-processing.

Table 5 displays the averages of all evaluation metrics for the various convolutional neural network architectures used in our experiments. We can see that fine-tuned ResNet152V2 outperforms the other

"Cervical Cancer Detection and Classification Using MRIs", I. Khoulqi and N. Idrissi

Table 5. The averages of evaluation metrics achieved by each classifier during training and validation using the FP dataset (FP_Dataset).

	Fold	Accuracy %	Recall %	Specificity %	Precision %	F1-measure%
Xception	1	61,53	86,88	25,58	60,50	57,25
	2	66,34	68,85	62,79	66,74	66,48
	3	75,96	88,52	58,13	76,29	75,19
	4	80,58	86,88	71,42	80,48	80,40
	5	81,55	95,08	61,90	82,97	80,75
	6	86,40	93,44	76,19	86,62	86,20
	7	90,29	93,44	85,71	90,28	90,25
	8	88,34	95,08	78,57	88,64	88,17
Average		**71,18**	**77,66**	**61,83**	**71,50**	**70,93**
Standard Deviation		**9,70**	**8,11**	**17,30**	**9,97**	**10,63**
VGG16	1	66,99	93,44	28,57	69,38	62,49
	2	83,49	85,24	80,95	83,56	83,52
	3	87,37	83,60	92,85	88,38	87,47
	4	92,23	91,80	92,85	92,35	92,25
	5	89,32	95,08	80,95	89,51	89,19
	6	93,20	90,16	97,61	93,73	93,24
	7	94,17	1	86,04	94,70	94,09
	8	92,23	96,66	86,04	92,39	92,16
Average		**87,37**	**79,62**	**80,73**	**88,0**	**86,80**
Standard Deviation		**8,37**	**30,01**	**20,48**	**7,78**	**9,75**
InceptionResNetV2	1	73,78	98,36	38,09	79,69	70,46
	2	85,43	90,16	78,57	85,39	85,34
	3	88,34	91,80	83,33	88,32	88,30
	4	93,20	90,16	97,61	93,73	93,24
	5	93,20	95,08	90,47	93,19	93,19
	6	92,23	98,36	83,33	92,67	92,11
	7	91,26	91,66	90,69	91,30	91,27
	8	94,17	98,33	88,37	94,37	94,12
Average		**88,95**	**94,23**	**81,30**	**89,83**	**88,50**
Standard Deviation		**6,34**	**3,48**	**17,21**	**4,77**	**7,35**
MobileNetV2	1	73,04	80,88	61,70	72,76	72,76
	2	82,60	89,70	72,34	82,63	82,38
	3	81,73	88,23	72,34	81,68	81,54
	4	82,60	86,76	76,59	82,53	82,54
	5	84,34	82,35	87,23	85,02	84,45
	6	91,30	92,64	89,36	91,30	91,30
	7	96,52	98,52	93,61	96,55	96,50
	8	93,04	97,01	87,50	93,17	92,99
Average		**85,64**	**89,51**	**80,08**	**85,70**	**85,55**
Standard Deviation		**7,06**	**5,93**	**10,27**	**7,14**	**7,14**
NASNetLarge	1	60	69,11	46,80	59,50	59,68
	2	65,21	55,88	78,72	69,38	65,26
	3	67,82	57,35	82,97	72,50	67,82
	4	78,26	94,11	55,31	79,94	77,06
	5	87,82	91,17	82,97	87,79	87,78
	6	86,95	94,11	76,59	87,24	86,75
	7	92,17	92,64	91,48	92,21	92,18
	8	94,78	97,05	91,48	94,80	94,76
Average		**79,12**	**81,42**	**75,78**	**80,41**	**78,91**
Standard Deviation		**12,45**	**16,47**	**15,26**	**11,56**	**12,49**
ResNet152V2	1	70,19	74,19	64,28	70,31	70,24
	2	91,34	90,16	93,02	91,57	91,38
	3	91,34	95,08	86,04	91,40	91,29
	4	91,34	100	79,06	92,45	91,13
	5	92,30	88,52	97,67	93,02	92,35
	6	94,17	98,36	88,09	94,37	94,12
	7	94,17	95,08	92,85	94,17	94,17
	8	98,05	100	95,23	98,11	98,05
Average		**90,36**	**92,67**	**87,03**	**90,675**	**90,34**
Standard Deviation		**7,91**	**8,03**	**10,20**	**7,95**	**7,89**

155

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 08, No. 02, June 2022.

proposed five models with 90.36% average accuracy. Data augmentation helps and facilitates the DCNNs to learn the underground feature without the impact of rotation and scale; complicated transforms are not better due to the introduction of noise during feature extraction that leads to disturbing the learning process. The proposed stacked models have performed the best out of the state-of-the-art pre-trained models. As a result, the proposed models have achieved accuracies of 99.56%, 98.70% and 99.56% for XGB, Logistic Regression and AdaBoost as meta-learners, respectively. The suggested model's validity is justified by the values of recall, specificity, precision, F-measure and AUC, as shown in Table 6.

Table 6. Model's validity.

	Accuracy	Recall	Specificity	Precision	F1-score	AUC
Xception	88,79	92,70	83,15	88,79	88,73	0,87
VGG16	95,25	98,54	90,52	95,37	95,22	0,94
InceptionResNetV2	93,53	100	84,21	94,17	93,42	0,92
MobileNetV2	94,82	97,81	90,52	94,90	94,79	0,94
NasNetLarge	94,39	95,62	92,63	94,39	94,39	0,94
ResNet152V2	95,25	98,54	90,52	95,37	95,22	0,94
XGB	**99,56**	**100**	**98,94**	**99,57**	**99,56**	**0,99**
Logistic Regression	**98,70**	**100**	**96,84**	**98,73**	**98,70**	**0,98**
AdaBoost	**99,56**	**99,27**	**100**	**99,57**	**99,56**	**0,99**

Fine-tuned deep convolutional neural network architectures will produce better results, because they require less computation time than training a DCNN from scratch, allowing it to converge faster.

The base learners required about ten minutes for training purposes. All of the pre-trained models used to identify our female pelvis MRIs achieved the highest accuracy value; nevertheless, given the importance of detecting cervical cancer at early stages, efforts can be made to improve the sensitivity, precision and accuracy scores. As a result, meta-learners are proposed. The proposed models utilizing stacking generalization outperform the base models with 99.56 %, 98.70 % and 99.56 % accuracy for XGB, LR and Adaboost, respectively, which is about 4% higher and an increased precision value, implying the correctness of the predicted results. The results show that the proposed stacking generalized approach yields a high specificity rate, signifying no false positive predictions. The system is more reliable when the specificity is high.

Table 7. Hyper-parameter values.

Hyperparameters	values
Learning rate	0.00005
Optimizer	Adam
Batch size	32
Epochs	5

During the training process, the various hyperparameters used in this work were fixed. With a learning rate of 0.00005 and Adam as the optimizer, the batch size and epochs were set to 32 and 5, respectively, as cited in Table 7.

Table 8. Comparative analysis between state-of-the-art and the proposed models.

State-of-the-art		Accuracy %	Precision%	Recall%	F1-score%	AUC
ML algorithms (DT, RF, XGB) [52]	DT : 93,33%	80	100	89	-	
	RF : 93,33%	100	75	86	-	
	XGB : 93,33%	0	0	0	-	
ML algorithms, Majority Vote [51]		94%	97%	97%	-	0,97
DenseNet121 and The Mutation-based Atom Search Optimization (MASO) algorithm [53]		98,38%	98,58%	99,3%	98,25%	-
Proposed Ensemble Model 1:XGB		**99,56**	**99,57**	**100**	**99,56**	**0,99**
Proposed Ensemble Model 2:Logistic Regression		**98,70**	**98,73**	**100**	**98,70**	**0,98**
Proposed Ensemble Model 3:AdaBoost		**99,56**	**99,57**	**99,27**	**99,56**	**0,99**

4.2.2 Comparative Analysis

The experimental results show that the suggested ensemble models provide a significant rapid and effective solution in the cervical cancer detection field utilizing MRIs of female pelvis. Table 8 compares the performance of existing models to that of the proposed models. The lack of data to train the models leads to models that are poor at generalization. To resolve this concern, we employed transfer learning, which reuses models trained on large datasets, such as ImageNet. We also attempted to make as minimum as possible false predictions. Considering the statistics presented in Table 6, the suggested ensemble models achieve better generalization and less false predictions by beating state-of-the-art models.

5. CONCLUSIONS

To sum up, this article presents two different axes to detect and classify cervical cancer using MRIs, the first based on image processing techniques passing through three essential steps: pre-processing, segmentation and classification of pelvic MRIs. To realize this objective, we propose also two different techniques of segmentation and a variety of descriptors and classifiers that show a high efficiency in extracting regions of interest (ROIs) confirmed due to the obtained values of the parameters of evaluation (ZSI and SSIM) that are close to 1, which means a better segmentation, followed by description and classification into benign or malignant. The experiments performed on our female pelvis dataset approve that the powerful approach of segmenting ROIs is based on KM-DRLSE and Random Forest (RF) for the classification process after using the forward selection to pick only the most informative descriptors. The second axe uses deep transfer learning techniques to build classification models for classifying cervical cancer. Data augmentation pre-processing was performed to enhance our female pelvis dataset to overcome overfitting issues and six pre-trained models are employed. The accuracy rates obtained from Model1, Model2 and Model3 with 99.56%, 98.70% and 99.56%, respectively, show that the proposed ensemble models are capable of detecting CC.

REFERENCES

[1]   C. Xavier, J. Levêque and D. Riethmuller,"Cancers Gynécologiques Pelviens," ISBN: 978-2-294-72937-9, Elsevier Masson, 2013.

[2]   E. Belglaiaa and C. Mougin, "Le cancer du col de l'utérus: état des lieux et prévention au Maroc," Bulletin du Cancer, vol. 106, no. 11, pp. 1008-1022, 2019.

[3]   American Cancer Society, "Cervical Cancer," [Online], Available: https://www.cancer.org/cancer/cervical-cancer/about/key-statistics.html, February 2017.

[4]   J. Bethanney, G. Umashankar, D. Sindu and N. Basilica, "Classification of Cervical Cancer from MRI Images Using Multiclass SVM Classifier," International Journal of Engineering and Technology (UAE), vol. 7, no. 2, DOI: 10.14419/ijet.v7i2.25.12351, 2018.

[5]   M. Arya, N. Mittal and G. Singh, "Cervical Cancer Detection Using Segmentation on Pap Smear Images," Proc. of the International Conference on Informatics and Analytics (ICIA-16), pp. 1-5, DOI: 10.1145/2980258.2980311, 2016.

[6]   P. Robert and A. Celine Kavida, "Classification of Microscopic Cervical Cancer Images Using Regional Features and HSI Model," IJITEE Journal, vol. 8, no. 8S, pp. 24-28, June 2019.

[7]   K. Sneha and C. Arunvinodh, "Cervical Cancer Detection and Classification Using Texture Analysis," Biomedical and Pharmacology Journal, vol. 9, no. 2, pp. 663-671, 2016.

[8]   T. A. Sajeena and A. S. Jereesh, "Automated Cervical Cancer Detection through RGVF Segmentation and SVM Classification," Proc. of the IEEE International Conference on Computing and Network Communications (CoCoNet), DOI: 10.1109/CoCoNet.2015.7411260, Trivandrum, India, 2015.

[9]   S. R. Salian and S. D. Sawarkar, "Melanoma Skin Lesion Classification Using Improved Efficientnetb3," Jordanian J. of Computers and Inf. Technology (JJCIT), vol. 8, no. 1, pp. 45-56, 2022.

[10]   T. G. Debelee, A. Gebreselasie, F. Schwenker, M. Amirian and D. Yohannes, "Classification of Mammograms Using Texture and CNN Based Extracted Features", Journal of Biomimetics, Biomaterials and Biomedical Engineering, vol. 42, pp. 79-97, 2019.

[11]   M.Sandler, A.Howard, M.Zhu, A.Zhmoginov and L.Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks", Proc. of the IEEE Conf. on Computer Vision and Pattern Recog, pp. 4510-4520, 2018.

[12]   H. A. Almubarak, R. J. Stanley, R. Long et al., "Convolutional Neural Network Based Localized Classification of Uterine Cervical Cancer Digital Histology Images," Procedia Computer Science, vol. 114, pp. 281–287, 2017.

[13]   G. Sun, S. Li, Y. Cao and F. Lang, "Cervical Cancer Diagnosis Based on Random Forest," International Journal of Performability Engineering, vol. 13, pp. 446–457, 2017.

[14]   A. Makris, I. Kontopoulos and K. Tserpes, "COVID-19 Detection from Chest X-Ray Images Using Deep Learning and Convolutional Neural Networks," Proc. of the 11th Hellenic Conference on Artificial Intelligence, DOI: 10.1101/2020.05.22.20110817, 2020.

[15]   C. Yuan, Y. Yao, B. Cheng et al., "The Application of Deep Learning Based Diagnostic System to Cervical Squamous Intraepithelial Lesions Recognition in Colposcopy Images," Scientific Reports, vol. 10, article ID: 11639, DOI: 10.1038/s41598-020-68252-3, 2020.

[16]   S. Ho, E. Bullitt and G. Gerig, "Level-set Evolution with Region Competition: Automatic 3-D Segmentation of Brain Tumors," International Conference on Pattern Recognition, vol. I, pp. 532-535, Chapel Hill, NC 27599, USA.

[17]   K. Karantzalos and D. Argialas, "A Region-based Level Set Segmentation for Automatic Detection of Man-made Objects from Aerial and Satellite Images," Photogrammetric Engineering & Remote Sensing, vol. 75, no. 6, pp. 667–677, June 2009.

[18]   J. Fan, D. K. Y. Yau, A. K. Elmagarmid and W. G. Aref, "Automatic Image Segmentation by Integrating Color-edge Extraction and Seeded Region Growing," IEEE Transactions on Image Processing, vol. 10, no. 10, pp. 1454-1466, DOI: 10.1109/83.951532, 2001.

[19]   M. Kocher and R. Leonardi, "Adaptive Region Growing Technique Using Polynomial Functions For Image Approximation," Signal Processing, vol. 11, no. 1, pp. 47-60, 1986.

[20]   S. Saladi and N. Amutha Prabha, "Analysis of Denoising Filters on MRI Brain Images," International Journal of Imaging Systems and Technology, vol. 27, no. 3, pp. 201-208, 2017.

[21]   C. Tomasi and R. Manduchi, "Bilateral Filtering for Gray and Color Images," Proc. of the IEEE 6 th Int. Conf. on Computer Vision (IEEE Cat. No.98CH36271), pp. 839-846, Bombay, India. 1998.

[22]   S. Paris and F. Durand, "Fast Approximation of the Bilateral Filter Using a Signal Processing Approach," Proc. of European Conf. on Computer Vision (ECCV 2006), vol. 3954, pp. 568-580, 2006.

[23]   B. K. J. Veronica, "An Effective Neural Network Model for Lung Nodule Detection in CT Images with Optimal Fuzzy Model," Multimedia Tools and Applications, vol. 79, pp. 14291–14311, 2020.

[24]   B. Nugroho, E. Y. Puspaningrum and A. Yuniarti, "Performance of Face Recognition with Pre-processing Techniques on Robust Regression Method," Int. J. of GEOMATE, vol. 15, no. 50, pp. 101-106, 2018.

[25]   I. Khoulqi and N. Idrissi, "Segmentation and Classification of Cervical Cancer," Proc. of the IEEE 6th Int. Conf. on Optimization and Applications (ICOA), pp. 1-7, Beni Mellal, Morocco, 2020.

[26]   A. Herbulot, Nonparametric Statistical Measurements for Image and Video Segmentation and Active Contour Minimization, Ph.D. Thesis, University of Nice-Sophia Antipolis, Nice, France, Oct. 2007.

[27]   Jiang Xin, R. Zhang and S. Nie, "Image Segmentation Based on Level Set Method," Physics Procedia, vol. 33, pp. 840-845, DOI: 10.1016/j.phpro.2012.05.143, 2012.

[28]   F. Chen, "Medical Image Segmentation Using Level Sets," Technical Report, University of Waterloo, Canada, pp. 1-8, 2008.

[29]   C. Li, C. Xu, C. Gui and M. D. Fox, "Distance Regularized Level Set Evolution and Its Application to Image Segmentation," IEEE Transactions on Image Processing, vol. 19, no. 12, pp. 3243-3254, 2010.

[30]   A. Khadidos, V. Sanchez and C.-T. Li, "Weighted Level Set Evolution Based on Local Edge Features for Medical Image Segmentation," IEEE Trans. on Image Process., vol. 26, no. 4, pp. 1979- 1991, 2017.

[31]   J. Bethanney and E. Roslin, "Classification and Detection of Skin Cancer using Hybrid Texture Features," Biomedicine, vol. 37, no. 2, pp.214-22, 2017.

[32]   Soumya M., Sneha K. and Arunvinodh C., "Cervical Cancer Detection and Classification Using Texture Analysis, "Biomedical and Pharmacology Journal, vol. 9, no. 2, 2016.

[33]   I. Khoulqi and N. Idrissi, "Breast Cancer Image Segmentation and Classification," Proc. of the 4th International Conference on Smart City Applications, pp. 1-9, DOI:10.1145/3368756.3369039, 2019.

[34]   S. Paris, "Le Multimédia et la Compression," Lavoisier, Hermes Science Publisher, 1, p. 208, Informatique, Jean-Charles Pomerol, 978-2-7462-2203-8.hal-00841603, 2009.

[35]   W. Lin, Z. Wu, L. Lin, A. Wen and J. Li, "An Ensemble Random Forest Algorithm for Insurance Big Data Analysis," IEEE Access, vol. 5, pp. 16568-16575, DOI: 10.1109/ACCESS.2017.2738069, 2017.

[36]   I. M. Nasser and S. S. Abu-Naser, "Predicting Tumor Category Using Artificial Neural Networks," International Journal of Academic Health and Medical Research (IJAHMR), vol. 3, no. 2, pp. 1-7, 2019.

[37]   R. V. K. Reddy et al., "Prediction of Heart Disease Using Decision Tree Approach," International Journal of Advanced Research in Computer Science and Software Engineering, vol. 6, no. 3, 2016.

[38]   L. E. Peterson, "K-nearest Neighbor," Scholarpedia, vol. 4, no. 2, Article ID: 1883, 2009.

[39]   C. Krauss, X. A. Do and N. Huck, "Deep Neural Networks, Gradient-boosted Trees, Random Forests: Statistical Arbitrage on the S&P 500, " European Journal of Operational Research, vol. 259, no. 2, pp. 689-702, 2017.

[40]   D. Theckedath and R. R. Sedamkar, "Detecting Affect States Using VGG16, ResNet50 and SEResNet50 Networks, "SN Computer Science, vol. 1, Article no. 79, pp. 1-7, 2020.

[41]   G. Sun, S. Li, Y. Cao and F. Lang, "Cervical Cancer Diagnosis Based on Random Forest," International Journal of Performability Engineering, vol. 13, no. 4, pp. 446–457, 2017.

[42]   D. M. Rouse and S. S. Hemami, "The Role of Edge Information to Estimate the Perceived Utility of Natural Images," Visual Communications Lab., School of Elect. and Comp. Eng., Cornell Uni., 2009.

[43]   J. W. Prescott, M. Pennell, T. M. Best et al., "An Automated Method to Segment the Femur for Osteoarthritis Research," Proc. of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 6364–6367, DOI: 10.1109/IEMBS.2009.5333257, 2009.

[44]   P. K. Malli and S. Nandyal, "Machine Learning Technique for Detection of Cervical Cancer Using KNN and Artificial Neural Network," IJETTCS, vol. 6, no. 4, pp. 145-149, 2017.

[45]   K. P. Chandran and U. V. Ratna Kumari, "Improving Cervical Cancer Classification on MR Images Using Texture Analysis and Probabilistic Neural Network," Int. J. of Engineering Sciences & Research Technology, vol. 4, pp. 3141-3145, 2015.

[46]   A. I. Naimi and L. B. Balzer, "Stacked Generalization: An Introduction to Super Learning," European Journal of Epidemiology, vol. 33, no. 5, pp. 459-464, 2018.

[47]   F. Chollet et al., "Xception: Deep Learning with Depthwise Separable Convolutions," Proc. of the IEEE Conf. on Comp. Vision and Pattern Recog., DOI: 10.1109/CVPR.2017.195, Honolulu, HI, USA, 2017.

[48]   T. Majeed et al., "Problems of Deploying CNN Transfer Learning to Detect Covid-19 from Chest Xrays," MedRxiv, DOI: 10.1101/2020.05.12.20098954, 2020.

[49]   J. Dekhtiar et al., "Deep Learning for Big Data Applications in CAD and PLM–Research Review, Opportunities and Case Study," Computers in Industry, vol. 100, pp. 227-243, 2018.

[50]   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 Recog. (CVPR), pp. 770-778, Las Vegas, NV, USA, 2016.

[51]   Q. M. Ilyas and M. Ahmad, "An Enhanced Ensemble Diagnosis of Cervical Cancer: A Pursuit of Machine Intelligence towards Sustainable Health," IEEE Access, vol. 9, pp. 12374-12388, 2021.

[52]   Akter, Laboni, Md Islam, Mabrook S. Al-Rakhami and Md Haque, "Prediction of Cervical Cancer from Behavior Risk Using Machine Learning Techniques," SN Computer Sci., vol. 2, ID 177, pp. 1-10, 2021.

[53]   B. Chitra and S. S. Kumar, "An Optimized Deep Learning Model Using Mutation-based Atom Search Optimization Algorithm for Cervical Cancer Detection," Soft Computing, vol. 25, no. 24, pp. 15363-15376, 2021.

ملخص البحث:

يُعدّ سرطان عُنق الرّحم ثاني أكثر الأمراض الخبيثة شيوعاً لدى النّساء بمعدّل وفيات قدره 60%، ويأتي في طليعة أسباب الوفاة على مستوى العالم. ويُمكن للتّحديد الدّقيق لمراحل المرض أن يُساعد خبراء الأشعّة والأطبّاء في تقديم علاج فعّال عبر الاستفادة من طُرق التّشخيص، مثل صُور الرّنين المغناطيسي. في هذه الورقة، نقترح طريقتيْن؛ تتمثّل الأولى في تقديم آليّ للكشف المبكّر عن سرطان عُنق الرّحم باستخدام تقنيات معالجة الصّور، وتحليل صُور الرّنين المغناطيسيّ لتحديد المرحلة التي وصل إليها الوَرَم والتّأثير الحقيقي له. وتمرّ عمليّة الكشف هذه في مراحل ثلاثٍ رئيسيّة (المعالجة الأوليّة لجعل صُور الرّنين المغناطيسي أسهل للمعالجة والتحليل؛ ثم التّجزئة من أجل استخلاص ما يُسمّى "المناطق ذات الاهتمام"، وفي الخطوة التالية يتم استخلاص فئتيْن من السّمات بناءً على طرقٍ إحصائية وطرق تحويل من أجل وصف صُور الرّنين المغناطيسي). وفي المرحلة النّهائية، تمّ تدريب خمسة مصنِّفات لتصنيف صُور الرّنين المغناطيسي الى صِنفيْن: (حميد وخبيث). أما الطّريقة الثّانية فتهدف الى زيادة أداء الشّبكة العصبية الالتفافيّة العميقة المدرَّبة مسبقاً بناءً على "تعلُّم التّحويل" المستخدَم لتصنيف مجموعات البيانات الخاصّة بالحَوْض لدى النّساء (FP) عن طريق استخدام طريقة التّكديس العامّة التي من شأنها أن توفّر مُصَنِّفاً أكثر فاعلية ومتانةً. وتعدّ زيادة البيانات إحدى طرق المعالجة الأوّلية التي تطبّق على صُور الرّنين المغناطيسي، كما تُستخدم طبقة إسقاطٍ لمنع الشّبكة من فَرْط التّهيئة في مجموعة البيانات الصّغيرة. وقد بينت نتائج التّجريب أنّ زيادة البيانات وتعميم التكديس يعدّان طريقة فعّالة لتحسين معدّل دقّة التّصنيف.

DEVELOPMENT OF ENSEMBLE MACHINE LEARNING MODEL TO IMPROVE COVID-19 OUTBREAK FORECASTING

Meaad Alrehaili and Fatmah Assiri

ABSTRACT

The world is currently facing the coronavirus disease 2019 (COVID-19 pandemic). Forecasting the progression of that pandemic is integral to planning the necessary next steps by governments and organizations. Recent studies have examined the factors that may impact COVID-19 forecasting and others have built models for predicting the numbers of active cases, recovered cases and deaths. The aim of this study was to improve the forecasting predictions by developing an ensemble machine-learning model that can be utilized in addition to the Naïve Bayes classifier, which is one of the simplest and fastest probabilistic classifiers. The first ensemble model combined gradient boosting and random forest classifiers and the second combined support vector machine and random-forest classifiers. The numbers of confirmed, recovered and death cases will be predicted for a period of 10 days. The results will be compared to the findings of previous studies. The results showed that the ensemble algorithm that combined gradient boosting and random-forest classifiers achieved the best performance, with 99% accuracy in all cases.

1. INTRODUCTION

The world is currently in the midst of a critical pandemic, the coronavirus disease 2019 (COVID-19 pandemic), which has spread throughout the world and is expected to continue doing so. By the end of 2019, there had been 7,000 deaths due to COVID-19 in 150 countries, prompting the World Health Organization (WHO) to declare the COVID-19 outbreak a global pandemic in March 2020 [1]. Previous research about pandemics allows for prediction of future pandemic cases or expansions, but the historical data must be reliable to ensure prediction accuracy [2].

Forecasting a pandemic's progression is extremely important for governmental and organizational actions, such as those within the fields of transportation, health-care and supplies. Prediction of the next phases of a pandemic will give the decision-making parties early notice of actions that they should undertake in order to minimize or even avoid catastrophes [3]. Additionally, successful forecasting will help control the situation by assisting the authorities in taking the right action at the right time to contain the crisis, thus preventing major losses. Many studies have been conducted on COVID-19 forecasting [3] and the existence of recently collected datasets and commitments of support from governments, health organizations and social parties enrich the opportunity to use machine-learning models to predict the progression of the pandemic.

A study used supervised machine-learning models to forecast pandemic development; classifying COVID-19 dataset using four classifiers or machine-learning models: linear regression (LR), support vector machine (SVM), exponential smoothing (ES), least absolute shrinkage and selection operator (LASSO) [4]. Then, the models were trained and evaluated using the R^2 score, adjusted R^2, mean absolute error (MAE), mean square error (MSE) and root mean square error (RMSE). Each classifier was evaluated separately; the results revealed that ES achieved the highest accuracy in forecasting. LR and LASSO also performed well in forecasting the numbers of deaths and confirmed cases. Previous studies trained machine-learning models for COVID-19 forecasting; however, a small dataset was utilized [4]. Thus, there is a pressing need for more accurate models that use larger dataset.

M. Alrehaili and F. Assiri are with College of Computer Science and Engineering, University of Jeddah, Jeddah, Saudi Arabia. Emails: malrehaili0006.stu@uj.edu.sa and fyassiri@uj.edu.sa

The contribution of this paper is to develop a more accurate early-forecast model for COVID-19 using ensemble machine-learning algorithms. Furthermore, we also conducted a comparative analysis based on four measurements (R^2 score, adjusted R^2, MAE, MSE and RMSE) that compared the performance of the proposed ensemble algorithms with the results of a similar study [4]. We used the same dataset that was used by [4], but we applied the Naïve Bayes classifier, the simplest and fastest probabilistic classifier [5], which can be used for COVID-19 forecasting by requiring several linear parameters for the number of features or predictors as a variable in a learning problem.

In addition, we propose two ensemble models for improving COVID-19 forecasting: the gradient boosting and random-forest (GBRF) ensemble model and the support vector machine and random-forest (SVM+RF) ensemble model. Following the approach used by [4], we made predictions regarding the new confirmed cases, recovered cases and deaths that would occur in the next 10 days. We also used the same evaluation methods to compare the results of the proposed ensemble methods to those of the previous work by computing the R^2 -score, adjusted R^2, MAE, MSE and RMSE [4]; this allowed to show which model has the highest degree of accuracy in COVID-19 forecasting.

The rest of the paper is organized as follows: Section 2 describes the literature review. The dataset description is given in Section 3 and the methodology is presented in Section 4. The results and discussion are presented in Section 5. The conclusion is presented in Section 6.

2. LITERATURE REVIEW

Accurate forecasting serves various crucial clinical purposes, particularly for health-based systems. Computer-aided clinical predictive models have been used in various areas, including for predicting the progression of different diseases. In this study, we applied different prediction models to build a predictive model for COVID-19. In a recent study, the researchers proposed a system for detecting COVID-19 movements and progression by forecasting cases based upon real-time data [6]. SVM was compared with seven other classifiers and achieved the highest accuracy (92.95%). The results of the SVM and k-nearest neighbors' classifiers were the same.

Machine learning has also been used for COVID-19 survival analysis and discharge time likelihood prediction [7]. Several machine-learning algorithms were used for these purposes, including gradient boosting, component-wise gradient boosting and SVM. The results indicated that the gradient boosting survival model is the best model for prediction of patient survival. However, the details of the dataset were not mentioned.

Random forest was also used for COVID-19 patient health prediction [8]. Different algorithms (i.e., decision tree, support vector, Gaussian Naive Bayes and boosted random forest), were used for this purpose and their performances were compared. The boosted random-forest algorithm was the best-performing model (94%) for patient health prediction. COVID-19 management and progression predictions were performed through the use of mathematical modeling and artificial intelligence [5]. The results of the Naive Bayes and other classifiers were compared by computing their respective prediction accuracy levels. The Naive Bayes classifier showed 99.4% accuracy, which was the highest. However, the different classifiers were tested on different datasets, which might have affected their accuracy levels.

Another use of the Naive Bayes classifier was proposed in [9]. A model for obtaining computed tomography images for predicting the progression of COVID-19 was implemented and multiple classifiers were used. The Naive Bayes classifier showed 92.15% accuracy in feature selection and its accuracy was enhanced to 96.07% when another dataset was used. The average Naive Bayes classifier accuracy was 94.11%, similar to that of the convolution neural network (CNN). No advantage of using the Naive Bayes classifier over the CNN classifier or *vice versa* was reported. Several studies have tried to solve the problems related to predicting the movement or progression of the COVID-19 pandemic, but the current models' robustness needs to be improved. In this study, we applied different machine-learning algorithms to address the prediction accuracy limitations of the previous studies.

Convolution Neural Networks (CNNs) have also been used for the diagnosis of COVID-19 based on the classification of chest X-ray images [10]. A total of 178 X-ray images were used, of which 136 images were for COVID-19 patients and the rest of non-infected people. The proposed CNN model was integrated incrementally, starting with a single layer, and adding a convolutional layer at each increment.

161

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 08, No. 02, June 2022.

The results exhibited 99.5% accuracy. In another study, the Facebook Prophet model predicted the number of future infections over 90 days, taking into account the peak dates of confirmed cases for 6 of the most affected countries in the world [11]. A comparative analysis of the use of machine-learning and soft-computation models in the prediction of COVID-19 was also conducted [12]. The results showed that multi-layered perceptron and adaptive network-based fuzzy inference systems are among the best ones.

To improve prediction accuracy, a hybrid machine-learning model that consists of an adaptive network-based fuzzy inference system (ANFIS) and a multi-layered perceptron-imperialist competitive algorithm was proposed to perform COVID-19 forecasts in Hungary [13]. The model predicted the number of infected cases and mortality rates within 9 days based on time-series data. To further improve the performance of the prediction models, in [14], a grey wolf optimizer and an artificial neural network were applied to the same data used [13] and the results were promising.

Another study applied supervised machine learning to perform sentiment analysis as a decision support tool to better manage the pandemic [15]. The contribution of this work lies in the features set identified by the authors. The results showed that extra tree classifiers performed best compared to other algorithms, with an accuracy of 93%. Most recent work utilized mobile sensors to collect users' vital signs such as temperature and coughing patterns, which was subsequently combined with information entered by the users through mobile applications. Next, all data were used by applying an Artificial Neural Network (ANN) as a symptom-prediction algorithm to predict the likelihood of a user having COVID-19 [16].

3. DATASET

For a fair comparison between the proposed ensemble algorithms and those developed by Rustam [4], we used the novel COVID-19 dataset obtained from Johns Hopkins University (JHU), which contains data beginning on January 22, 2020 and is updated daily. The data is sourced from governments, national agencies across the world and the WHO [17], [24].

The number of global confirmed cases when the study was conducted was 10,853, 589 and the number of global deaths was 2,393,707. Figure 1 shows the average number of daily deaths and daily confirmed cases. The aforementioned dataset was accessed from the COVID-19 Data Repository of the Centre for Systems Science and Engineering at JHU. The data features include the state, region, date, number of confirmed cases, death cases and number of recovered cases. To meet the needs of this study, we further pre-processed the dataset.

Below are the descriptions of the features (attributes) that were used in this study [17]:

- Confirmed cases: The counts included the reported confirmed and probable cases.
- Deaths: The counts included the reported confirmed and probable cases.
- Recovered cases: Estimates based on local media reports and on state and local-government reports were considered where available; thus, this may be substantially lower than the true number.

4. METHODOLOGY

In this study, we used the Naive Bayes classifier, because it is one of the simplest and fastest classifiers, especially in the training stage [5]. We also utilized two ensemble models, GBRF and SVM+RM. SVM and random-forest machine-learning models complement each other; random forest computes the probability of belongings to a class, while SVM computes the distance to the boundary. Random forest can also complement gradient boosting, which is sensitive to noise and can cause overfitting [20].

Voting is a technique used to combine the results of many classifiers. There are three types of voting: unanimous, majority and plurality voting. In *unanimous voting,* all classifiers agree on a final decision. Majority voting makes the decision based on the number of voters; if one half or more of the votes go for one option, then it gets selected. In *plurality voting*, if most votes go to one option, it is selected as the final decision. In this study, we combine classifiers using majority voting, as it has been the most used one in prior studies.

The *Naive Bayes classifier* assumes that the classes' features are not related to each other, and it is not

affected by the classification assumption. It only requires a small training dataset to estimate the means and variances needed for classification [21]. Gradient boosting is a machine-learning model that generates a forecasting model in the form of an ensemble of weak-prediction models to increase the prediction performance. Gradient boosting is used for regression and classification problems [7].

Random forest is a common machine-learning method for developing prediction models in many research settings. To minimize the burden of data collection and to improve its efficiency, the random-forest model can be used as a prediction model to decrease the number of variables required to achieve a prediction [23]. Equation (1) presents RF regression model:

$$h(x) = \frac{1}{P}\sum_{n}^{p=1} h(y, \lambda P) \tag{1}$$

SVM is a statistical classifier that is used for linear and non-linear pattern classification [22]. The data is converted into high-dimension representations *via* non-linear mapping and SVM searches the new representations for the most appropriate data classification. SVM classifies data by increasing the margins of the classes; at the same time, it decreases the classification errors [21].

To predict the total number of people that might be affected in terms of new confirmed COVID-19 cases, deaths and expected recoveries for the upcoming 10 days, the Naive Bayes classifier and the ensemble models were trained using a dataset spanning from January to March 2020 [24]. The size of the training dataset was 66 days and that of the testing dataset was 10 days, following the approaches used in [4]. We evaluated the performances of the learning models in terms of R^2 score, adjusted R^2, MAE, MSE and RMSE, which are commonly used in the evaluation of predictive problems.

The R^2 score is only useful for simple linear regression. When using multiple linear regression, the value of the R^2 score grows as the number of independent variables increases, even if the independent variable is small. Adjusted R^2, on the other hand, increases only when the independent variable is significant and impacts the dependent variable. Equations (2) and (3) present the evaluation of R^2 and adjusted R^2, respectively:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{2}$$

$$Adjusted\ R^2 = 1 - \frac{(1-R^2)(N-1)}{N-p-1} \tag{3}$$

Mean Absolute Error (MAE) measures the differences between target values and predicted values, as shown in Equation (4):

$$MAE = \frac{\sum_{i=1}^{n}|X_i - \hat{X}_i|}{n} \tag{4}$$

Mean Square Error (MSE) takes the square of the differences between the actual and predicted values. This removes any negative values, as shown in Equation (5):

$$MSE = \frac{\sum_{i=1}^{n}(X_i - \hat{X}_i)^2}{n} \tag{5}$$

Root Mean Square Error (RMSE): detects the error rate from the regression model and compares the error size against the size of the target value. Equation (6) presents the evaluation of RMSE:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \hat{X}_i)^2}{n}} \tag{6}$$

Figure 2 shows the study's overall methodology.

5. RESULTS AND DISCUSSION

This evaluation study was designed to answer the following research questions:

- RQ1 (accuracy of the Naive Bayes classifier): What is the degree of accuracy of the Naive Bayes classifier?
- RQ2 (accuracy of the ensemble classifiers): What are the degrees of accuracy of the GBRF and SVM+RF ensemble classifiers?

This paper aims to build a predictive model using machine-learning algorithms for potential prediction of COVID-19 cases. The analysis provides details on regular estimates for the total number of confirmed, recovered and death cases around the world. The total of death and confirmed cases increased

163

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 08, No. 02, June 2022.

daily; this is obviously worrying. The following sub-sections discuss the results of the proposed models in terms of new infected, recovered and death cases.
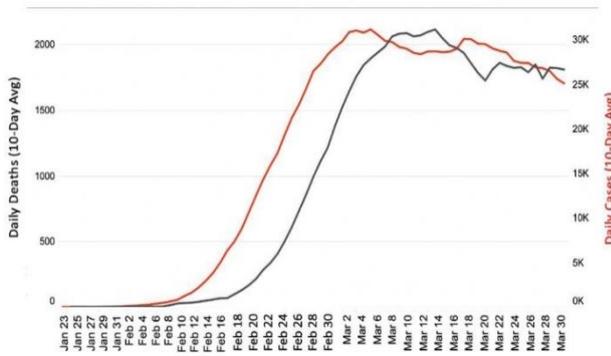
Figure 1. Daily deaths and daily confirmed cases (10-day average).
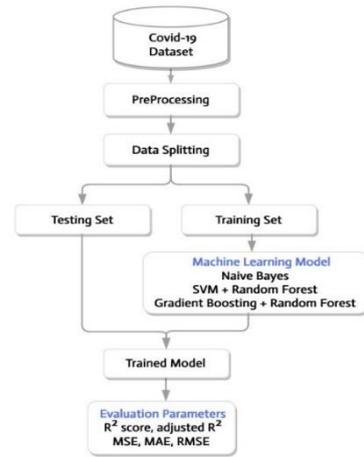
Figure 2. Proposed methodology workflow.

5.1 Future Forecasting of New Infections

The results of the Naive Bayes classifier for the number of cases of COVID-19 showed that the predicted number of cases was lower than the actual number of cases. As the attempted prediction period grew, the gap between the predicted and actual values increased. Figure 3 shows the Naive Bayes classifier predictions. The results of the SVM+RF ensemble model that was used to predict the number of new confirmed cases of COVID-19 showed that the predicted number of cases did not match the number of actual cases. The gap between the predicted values and the actual values increased with the number of upcoming days. Figure 4 shows the SVM+RF ensemble model predictions.
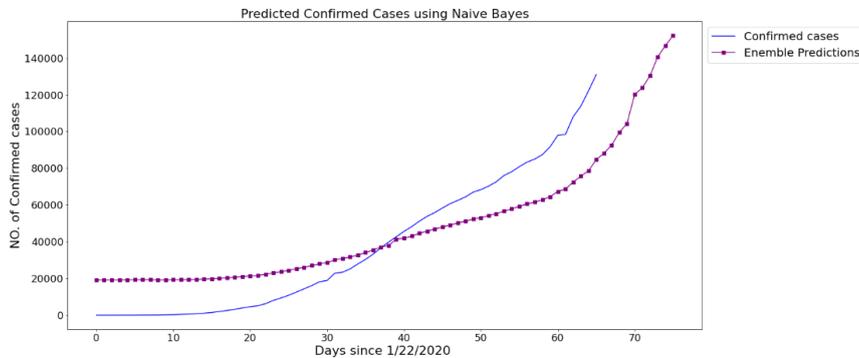
Figure 3. New infected cases for the upcoming 10 days using Naive Bayes.
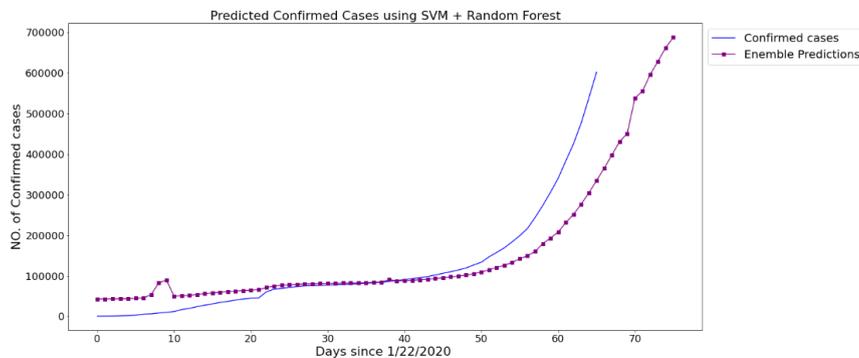
Figure 4. New infected cases for the upcoming 10 days using SVM and random forest.

The results of the GBRF ensemble model, which was used to predict the number of confirmed cases, showed that the predicted values matched the number of actual cases. Figure 5 shows the GBRF ensemble model predictions. Table 1 shows the results of the Naive Bayes classifier and the two

proposed ensemble models. GBRF gives the best results when predicting the newly infected cases for the upcoming 10 days. In contrast, Naïve Bayes and SVM+RF performed poorly.
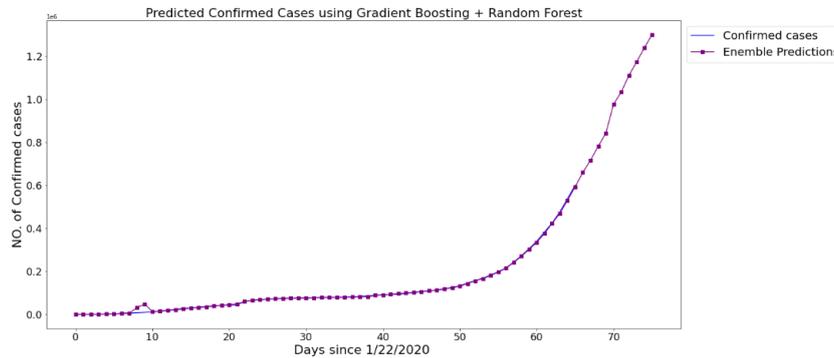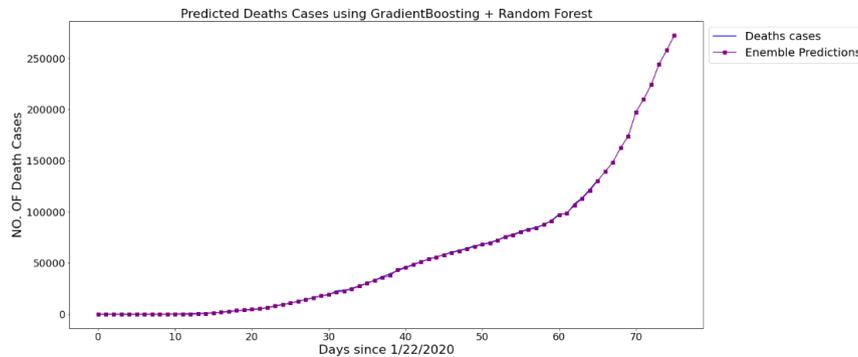

Figure 5. New infected cases for the upcoming 10 days using gradient boosting and random forest.

Table 1. Model performance of future forecasting for newly infected cases.

Models	R² score	Adjusted R²	MSE	MAE	RMSE
Naive Bayes	0.71	0.70	4628031693.91	39781.21	68029.64
SVM+RF	0.68	0.67	34836665919.42	98018.14	186645.83
GBRF	0.99	0.99	69138158.83	4092.88	8314.94

5.2 Future Forecasting of Recovered Cases

The results of the Naive Bayes classifier for recovered cases showed that the predicted number of recovered cases was lower than the actual number. With an increase in the number of upcoming days, the gap between the predicted and the actual values increased. Figure 6 shows the Naive Bayes classifier predictions.
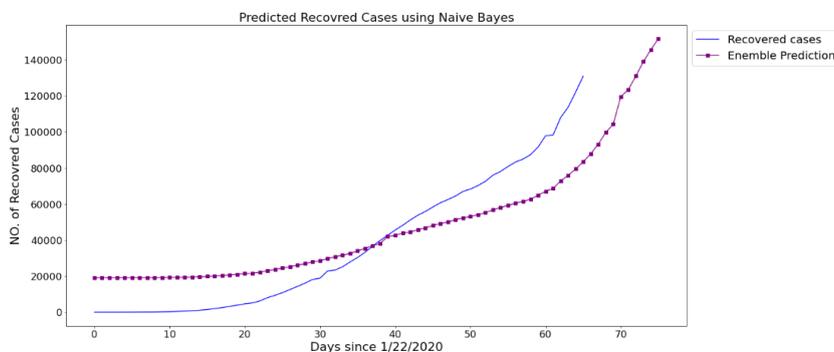
Figure 6. Recovery rate prediction for the upcoming 10 days using Naive Bayes.

The results of the SVM+RF ensemble model prediction of the number of recovered cases did not match the actual number of recovered cases. The predicted values were less than the actual cases and after five more days, the predicted number of cases became more than the number of actual cases. Figure 7 shows the SVM+RF ensemble model predictions.
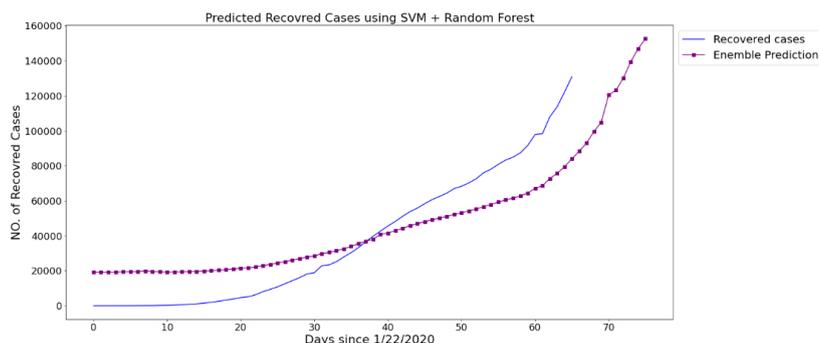
Figure 7. Recovery rate prediction for the upcoming 10 days using support vector machine and random forest.

The results of the GBRF ensemble model prediction for the number of recovered cases did not match the actual number of cases. Figure 8 shows the GBRF ensemble model predictions. The performance results of the models when predicting the number of recovered cases is shown in Table 2. GBRF gave the best results when predicting the recovered cases for the upcoming 10 days, while Naive Bayes and SVM+RF gave approximately the same results.
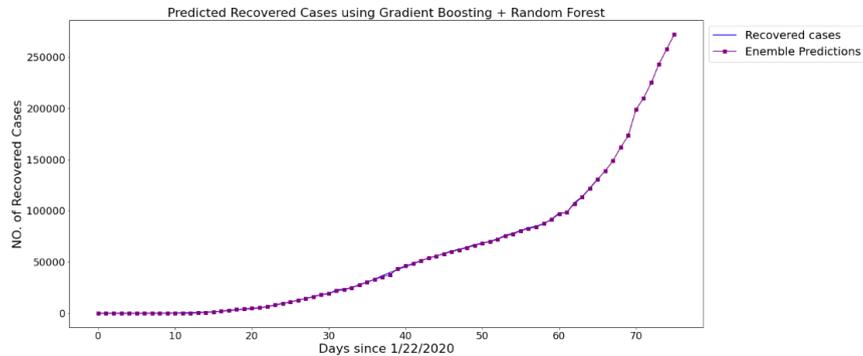
Figure 8. Recovery rate prediction for the upcoming 10 days using gradient boosting and random forest.

Table 2. Model performance for future forecasting for recovered cases.

Models	R² score	Adjusted R²	MSE	MAE	RMSE
Naive Bayes	0.72	0.73	4628031693.91	41781.21	68029.64
SVM+RF	0.71	0.70	1335741927.85	25459.42	36547.8
GBRF	0.99	0.99	1682739.6	711.0	1297.2

5.3 Future Forecasting of the COVID-19 Death Rate

The results of the Naive Bayes classifier prediction of the number of COVID-19 deaths showed that the number of predicted cases was lower than the actual number. With an increase in the number of upcoming days, the gap between the predicted and actual values increased. Figure 9 shows the Naive Bayes classifier predictions. The SVM+RF ensemble model predictions for the number of deaths did not match the actual number of deaths. Figure 10 shows the SVM+RF ensemble model predictions.
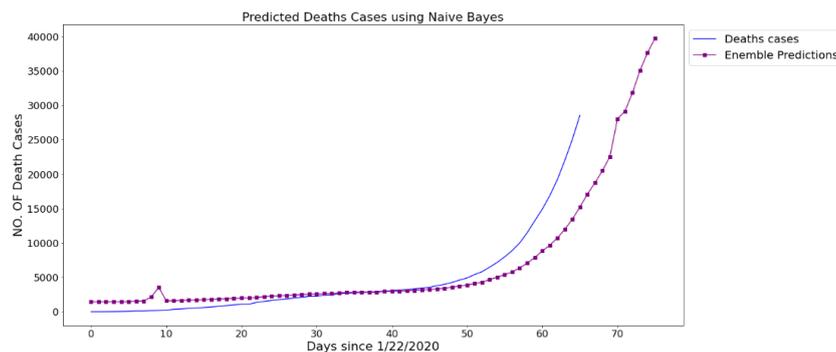
Figure 9. Death prediction for the upcoming 10 days using Naive Bayes.
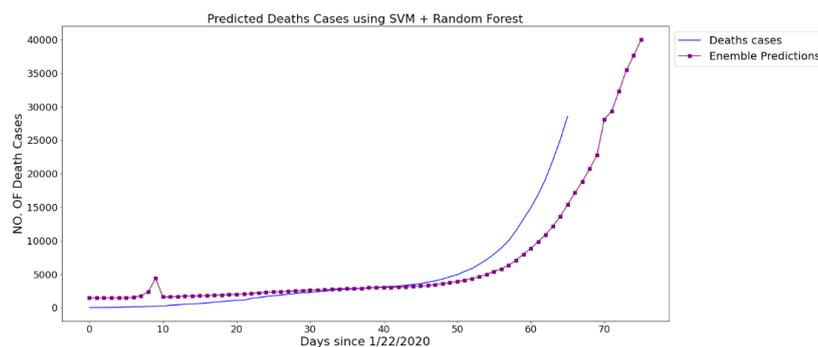
Figure 10. Death prediction for the upcoming 10 days using support vector machine and random forest.

The results of the GBRF predictions of the number of deaths showed that the predicted value was almost the same as the number of actual deaths. A gap appeared only in the eighth and ninth days, as shown in Figure 11 The death rate prediction performance results are shown in Table 3. GBRF gave the best results when predicting death cases for the upcoming 10 days. In contrast, Naive Bayes and SVM+RF performed poorly.

Figure 11. Death prediction for the upcoming 10 days using gradient boosting and random forest.

Table 3. Model performance of future forecasting for death cases.

Models	R^2 score	Adjusted R^2	MSE	MAE	RMSE
Naive Bayes	0.72	0.69	10833945.26	41181.5	68029.62
SVM+RF	0.69	0.68	106804953.79	5032.5	10334.65
GBRF	0.99	0.99	200297.36	215.2	447.55

To summarize our findings, the GBRF ensemble model reached the highest accuracy (99%) in all the cases, performing better than the Naive Bayes classifier and the SVM+RF ensemble model.

5.4 Model Performance within 10-Day Prediction Intervals

We compared the results of the applied models (i.e., Naive Bayes classifier, SVM+RF ensemble model and GBRF ensemble model) with the results of other models (i.e., linear regression (LR), support vector machine (SVM), exponential smoothing (ES), least absolute shrinkage and selection operator (LASSO)) tested in [4]. Table 4 shows the comparison results between the applied ensemble machine-learning models and other models tested in [4]. We then compared the two best models, GBRF and ES in three aspects:

- Confirmed-cases Prediction: GBRF gave higher R^2 score and higher adjusted R^2 values, which means that GBPR is better than SE at predicting the number of newly confirmed cases. Also, the MSE, RMSE and MAE results for GBRF were less than those of ES.
- Recovered-cases Prediction: The R^2 for GBRF was better than that for ES. The adjusted R^2 value was the same for GBRF and ES. MSE, RMSE and MAE values for GBRF were less than those for ES.
- Number-of-deaths Prediction: The results were similar to the results of the confirmed-cases prediction. The results of GBRF were better than those of ES in all evaluation measurements used in this work.

Table 4. Comparison between the proposed ensemble machine-learning models and the models for future forecasting in [4].

Model	Evaluation	Confirmed	Recovered	Death
Naïve Bayes	R^2 score	0.71	0.72	0.72
	Adjusted R^2	0.70	0.73	0.696
	MSE	4628031.7 k	4628031.7 k	10834 k
	MAE	39781.21	41781.21	1900.88
	RMSE	68029.64	68029.64	3291.5

167

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 08, No. 02, June 2022.

SVM+RF	R^2 score	0.68	0.71	0.69
	Adjusted R^2	0.67	0.70	0.68
	MSE	34836666 k	1335742 k	106805 k
	MAE	98018.14	25459.42	5032.5
	RMSE	186645.83	36547.8	10334.65
GBRF	R^2 score	0.999	0.999	0.999
	Adjusted R^2	0.99	0.99	0.99
	MSE	69138 k	1683 k	200 k
	MAE	4092.88	711.0	215.2
	RMSE	8314.94	1297.2	447.55
ES	R^2 score	0.98	0.99	0.98
	Adjusted R^2	0.97	0.99963	0.97
	MSE	0.67	34836665919.42	98018.14
	MAE	8867.43	1827.85	406.08
	RMSE	16828.58	2243.48	813.77
LR	R^2 score	0.83	0.39	0.96
	Adjusted R^2	0.79	0.21	0.95
	MSE	1472986 k	480922 K	840240.11
	MAE	30279.55	30705.27	723.11
	RMSE	38390.51	21929.95	916.64
LASSO	R^2 score	0.98	0.29	0.85
	Adjusted R^2	0.97	0.08	0.81
	MSE	234489 k	1462144 k	3244066.79 k
	MAE	11693.97	30705.27	1430.29
	RMSE	15322.11	38237.99	1801.12
SVM	R^2 score	0.59	0.24	0.53
	Adjusted R^2	0.02	0.99963	0.39
	MSE	5760890 k	13121148 k	160162 k
	MAE	60177.9	106739.82	3129.74
	RMSE	75911.28	114547.58	4002.02

To summarize our findings, GBRF performed best in the current forecasting domain given the nature and size of the dataset, followed by ES. LR and LASSO performed fairly well in forecasting death rates and newly confirmed cases. The Naive Bayes classifier and the SVM+RF ensemble model showed approximately the same degree of accuracy. SVM produced poor results in all the scenarios.

6. CONCLUSION

Forecasting the movement and progression of a pandemic facilitates governmental or organizational actions needed to contain that pandemic. During the current COVID-19 pandemic, forecasting is essential to prevent high numbers of active cases and deaths. Machine learning-based prediction models have been proposed for predicting the risk of COVID-19 outbreak. In this study, we used the Naive Bayes classifier and two ensemble algorithms: GBRF and SVM+RF. We compared these models with the prediction models tested in [4].

Our results showed that the best performance when forecasting new infections, recovered cases and deaths was achieved by the GBRF ensemble model (99%). This is an improved performance over ES, which reached 98% in [4]. The Naive Bayes classifier and SVM+RF ensemble model showed approximately the same performance, reaching 71% and 68%, respectively. SVM performed poorly in all scenarios. These results could be due to the nature of the significance improvement of ensemble models, in which two or more algorithms complement each other to provide better results [25]. Deep-learning algorithms have been known for their advantages over machine-learning models and we will consider them in our future work.

"Development of Ensemble Machine Learning Model to Improve COVID-19 Outbreak Forecasting", M. Alrehaili and F. Assiri.

ACKNOWLEDGMENT

REFERENCES

[1]   Centers for Disease Control and Prevention, "Severe Outcomes among Patients with Coronavirus Disease 2019 (COVID-19) : United States," MMWR Morb. Mortal. Wkly. Rep., vol. 69, no. 12, pp. 343-346, February 12–March 16, 2020.

[2]   I. M. Hall et al., "Real-time Epidemic Forecasting for Pandemic Influenza," Epidemiology and Infection, vol. 135, no. 3, pp. 372-385, 2007.

[3]   K. Sarkar, S. Khajanchi and J. Nieto, "Modeling and Forecasting the COVID-19 Pandemic in India," Chaos, Solitons & Fractals, vol. 139, DOI : 10.1016/j.chaos.2020.110049, 2020.

[4]   F. Rustam, A. Reshi, A. Mehmood, S. Ullah, B. Won On et al., "COVID-19 Future Forecasting Using Supervized Machine Learning Models," IEEE Access, vol. 8, pp. 101489–101499, 2020.

[5]   Y. Mohamadou, A. Halidou and P. T. Kapen, "A Review of Mathematical Modeling, Artificial Intelligence and Datasets Used in the Study, Prediction and Management of COVID-19," Applied Intelligence, vol. 50, pp. 3913–3925, 2020.

[6]   M. Otooma, N. Otoumb, M. A. Alzubaidi, Y. Etoom and R. Banihani, "An IoT-based Framework for Early Identification and Monitoring of COVID-19 Cases," Biomedical Signal Processing and Control, vol. 62, DOI : 10.1016/j.bspc.2020.102149, 2020.

[7]   M. Nemati, J. Ansary and N. Nemati, "Machine-learning Approaches in COVID-19 Survival Analysis and Discharge-time Likelihood Prediction Using Clinical Data," Patterns, vol. 1, no. 5, 2020.

[8]   C. Iwendi, A. K. Bashir, A. Peshkar et al., "COVID-19 Patient Health Prediction Using Boosted Random Forest Algorithm," Frontiers in Public Health, vol. 357, DOI : 10.3389/fpubh.2020.00357, 2020.

[9]   A. Farid, G. Selim and H. Khater, "Novel Approach of CT Images Feature Analysis and Prediction to Screen for Corona Virus Disease (COVID-19)," International Journal of Scientific and Engineering Research, vol. 11, no. 3, DOI:10.14299/ijser.2020.03.02, 2020.

[10]   A. A. Reshi et al., "An Efficient CNN Model for COVID-19 Disease Detection Based on X-ray Image Classification," Complexity, vol. 2021, DOI: 10.1155/2021/6621607, 2021.

[11]   S. Dash et al., "Intelligent Computing on Time-series Data Analysis and Prediction of COVID-19 Pandemic," Pattern Recognition Letters, vol. 151, pp. 69-75, 2021.

[12]   S. F. Ardabili et al., "COVID-19 Outbreak Prediction with Machine Learning," Algorithms, vol. 13, no. 10, DOI: 10.3390/a13100249, 2020.

[13]   G. Pinter et al., "COVID-19 Pandemic Prediction for Hungary; A Hybrid Machine Learning Approach," Mathematics, vol.8, no.6, DOI: 10.3390/math8060890, 2020.

[14]   S. Ardabili et al., "Coronavirus Disease (COVID-19) Global Prediction Using Hybrid Artificial Intelligence Method of ANN Trained with Grey Wolf Optimizer," Proc. of the 3rd IEEE International Conference and Workshop in Óbuda on Electrical and Power Engineering (CANDO-EPE), DOI: 10.1109/CANDO-EPE51100.2020.9337757, Budapest, Hungary, 2020.

[15]   F. Rustam et al., "A Performance Comparison of Supervized Machine Learning Models for COVID-19 Tweets Sentiment Analysis," Plos One, vol. 16, no. 2, DOI: 10.1371/journal.pone.0245909, 2021.

[16]   H. Khaloufi et al., "Deep Learning Based Early Detection Framework for Preliminary Diagnosis of COVID-19 via Onboard Smartphone Sensors," Sensors, vol. 21, no. 20, DOI: 10.3390/s21206853, 2021.

[17]   J. H. C. R. Center, Johns Hopkins Coronavirus Resource Center, [Online], Available: https://coronavirus.jhu.edu/.

[18]   The COVID Tracking Project, [Online], Available: https://covidtracking.com/data.

[19]   U. N. Dulhare, "Prediction System for Heart Disease Using Naive Bayes and Particle Swarm Optimization," Biomedical Research, vol. 29, pp. 2646–2649, DOI:10.4066/biomedicalresearch.29-18-620, 2018.

[20]   W. Zhang, C. Wu, H. Zhong, Y. Li and L. Wang, "Prediction of Undrained Shear Strength Using Extreme Gradient Boosting and Random Forest Based on Bayesian Optimization," Geoscience Frontiers, vol. 12, no. 1, pp. 469–477, 2021.

[21]   V. Mohan, "Liver Disease Prediction Using SVM and Naive Bayes Algorithms," International Journal of Science, Engineering and Technology Research (IJSETR), vol. 4, no. 4, pp. 816–820, 2015.

[22]   M. Loey, G. Manogaran, M. Taha and N. Khalifa, "A Hybrid Deep Transfer Learning Model with Machine Learning Methods for Face Mask Detection in the Era of the COVID-19 Pandemic," Measurement, vol. 167, DOI: 10.1016/j.measurement.2020.108288, 2020.

[23]   J. L. Speiser, M. E. Miller, J. Tooze and E. Ip, "A Comparison of Random Forest Variable Selection Methods for Classification Prediction Modeling," Expert Systems with Applications, vol. 134, pp. 93–101, DOI: 10.1016/j.eswa.2019.05.028, 2019.

[24]   J. Hopkins, Johns Hopkins University Data Repository, [Online], Available : https://github.com/CSSEGISandData.

[25]   N. Singhal et al., "Comparative Study of Machine Learning and Deep Learning Algorithm for Face Recognition," Jordanian Journal of Computers and Information Technology (JJCIT), vol. 7, no. 3, pp. 313-325, Sep. 2021.

ملخص البحث:

تهـدف هـذه الدراسـة الـى تحسـين توقُّـع انتشـار جائحـة فيـروس كورونـا عـن طريـق تطـوير نمـوذج موحَّـد مـن نمـاذج تعلُّـم الآلـة مـن الممكـن الاسـتفادة منـه بإضـافته الـى مصنِّـف مـن نـوع (Bayes Naïve) الـذي يُعـدّ واحـداً مـن أبسـط المصنِّفات الاحتماليـة وأسـرعها. النّمـوذج الأول يجمـع بـين تعزيـز الميْـل و مصنِّفات الغابـة العشـوائية (RF) بينمـا يجمـع الثـاني بـين آلـة متّجهـات الـدّعم (SVM) ومصنِّفات الغابـة العشـوائية. وسـيتم توقُّـع حـالات الإصـابة المؤكَّـدة، وحـالات الشّـفاء، وحـالات الوفـاة بسـبب الجائحـة لمـدّة عشـرة أيام. كذلك ستجري مقارنة نتائج هذه الدّراسة مع نتائج دراسات سابقة.

وبيّنـت النتـائج أنّ النّمـوذج الـذي يجمـع بـين تعزيـز الميْـل ومصنِّفات الغابـة العشـوائية حقَّق الأداء الأفضل بدقّة بلغت 99% في جميع الحالات.

BANDWIDTH ENHANCEMENTS AND SIZE REDUCTION OF 3 dB PATCH COUPLER WITH 45° OUTPUT PHASE DIFFERENCE FOR 5G BEAMFORMING NETWORKS

Nazleen S. M. Suhaimi, Anthony N. Uwaechia and Nor M. Mahyuddin

ABSTRACT

In this article, a single-layered 3 dB/45° coupler (Design D) is proposed for fifth-generation (5G) beamforming networks using cross-slotted patch topology, dumbbell-shaped slots, loaded stubs, notches and rectangular ground slots. The proposed Design D coupler is capable of eliminating the need for additional 45° phase shifters in the beamforming networks such as Butler matrix, which provides the main contribution in this work, especially in size reduction and bandwidth enhancements. The fractional bandwidths of 28.90%, 39.14% and 35.91% for -10 dB of $|S_{11}|$, -3 dB ± 1 dB of $|S_{31}|$ and 5° phase imbalance of output phase difference are achieved by the proposed Design D coupler. The bandwidth enhancements of -3 dB ± 1 dB coupling coefficient, S_{31} and 45° ± 5° output phase difference for the proposed Design D coupler are 22.52% and 23.14% compared to Design A coupler, respectively. The bandwidth of 45° ± 5° output phase difference is increased by 16.3% owing to the presence of rectangular ground slots in Design D compared to Design C coupler. The patch size of the proposed Design D coupler is 0.22 λ_g x 0.23 λ_g. The electrical size of the proposed Design D coupler is reduced by 45.72% compared to Design A coupler.

KEYWORDS

5G beamforming networks, Patch coupler, Output phase difference, Bandwidth enhancement, Size reduction.

1. INTRODUCTION

As a pioneer in fifth-generation (5G) wireless communication, the 6 GHz band is currently allocated in many countries [1]-[2]. A good trade-off between equipment size, network coverage and ubiquitous high-speed connectivity in all locations becomes a remarkable challenge for researchers and engineers involved in the development of radio frequency (RF) and microwave components, antenna design and network planning.

Passive microwave couplers are the main components found in beamforming networks such as Butler matrix [3]-[7], Blass matrix [8]-[11] and Nolen matrix [12]-[13]. Branch-line couplers are extensively utilized in beamforming networks to provide an equal 3 dB power split with a standard 90° output phase difference within the desired frequency band [14]. Nevertheless, the traditional branch-line couplers provide large size and narrow bandwidth. The restricted standard 90° output phase difference between the output signals of the traditional branch-line coupler further restricts its applications. The reported improvements in the branch-line coupler performance mainly focus on multiband [15]-[19], arbitrary power division ratio [20]-[21], broadband [22]-[24], compact size [25] and integration with additional phase shifter [26]-[27]. The traditional couplers can provide output phase differences of 0°, 90° and 180°, but the non-standard phase difference such as 30°, 45° or 150° could not be realized without additional phase shifters. Nowadays, the non-standard phase characteristic coupler is being potentially highly demanded in smart antenna systems, measurement tools and power amplifiers. The most straightforward method to solve this problem is to propose additional phase shifters, but they contribute major drawbacks, such as large circuit size and inevitable large insertion loss between main line and reference line. The unequal insertion loss is further aggravated through the final stage of a feed network causing intolerable magnitude fluctuations among other output ports. Although bandwidths of the phase shifters in [28]-[30] are enhanced, these works are undesirable for low-cost applications due to high fabrication complexities.

As a step towards 5G technology, there is extremely high demand for a wideband operation, compact in size and low cost in modern microwave circuits. Therefore, the coupler structure which can provide any non-standard output phase difference without any additional phase shifter circuit quickly draws

N. S. M. Suhaimi, A. N. Uwaechia and N. M. Mahyuddin are with School of Electrical and Electronic Engineering, Universiti Sains Malaysia, Penang, Malaysia. Emails: nsyahira@student.usm.my, anthony_uwaechia@usm.my and eemnmuzlifah@usm.my

attention. Furthermore, an effective approach to achieve impedance matching and phase stabilization is desired. The previous branch-line couplers have output phase difference of 90° and other output phase differences could only be realized using additional phase shifters. This increased circuit size and cost as well as degraded the overall performances owing to interconnecting mismatch losses.

Nowadays, Butler matrix is one of the most popular realizations of the beamforming network. Nevertheless, the required output phase differences found in the Butler matrix are 45° and 135°, which cannot be realized with the traditional power divider or branch-line coupler without using additional phase shifters. In order to circumvent the arising issues, the proposed 3 dB patch coupler with 45° output phase difference is capable of eliminating the need for additional 45° phase shifters, which provides the main contribution in this work, especially in size reduction and bandwidth enhancements of S-parameter responses and output phase difference.

In this work, a single-layered 3 dB patch coupler with 45° output phase difference is proposed without using any phase shifter. The bandwidth performance of the desired phase difference for the proposed 3 dB/45° patch coupler can be controlled by selecting the suitable dimensional lengths and widths of the dumbbell-shaped slots, whilst the bandwidth performances of the reflection and coupling coefficients are controlled by the dimensional lengths and widths of the cross slots and loaded stubs. By comparing the couplers in [31-33], the bandwidth performances of reflection and coupling coefficients are controlled by impedances of transmission lines, whilst the output phase difference is controlled by the electrical lengths of the transmission lines.

Conventionally, the traditional 4×4 Butler matrix comprises four traditional 3 dB/90° branch-line couplers, two traditional 0 dB crossovers, two 0° phase shifters and two 45° phase shifters. Meanwhile, the physical size of 4×4 Butler matrix can be reduced using two proposed 3 dB/90° patch couplers (Design B) with a size reduction of 26.32% and two proposed 3 dB/45° patch couplers (Design D) with a size reduction of 45.72% compared to their traditional designs. The Butler matrix can be realized without using any phase shifters and crossovers as reported in [34], thereby making it remarkably of low transmission loss, low insertion loss, less number of components and more compact compared to the traditional Butler matrix. The patch topology is adopted from [35], where the inductive loading effect of cross slots results in a significant size reduction. The parametric analyses are carried out on the designed couplers to analyse the effects of cross slots' and stubs' length variations on the S-parameter bandwidth performances. Moreover, the power division ratio and output phase difference performances are investigated and analyzed. The 90° output phase difference is reduced to 45° by introducing the dumbbell-shaped slots at the end of the cross slots as well as optimizing the lengths and widths of the cross slots and rectangular stubs. Meanwhile, the chamfering corners are optimized to readjust the impedance matching to 50 Ω. Notches are placed along each length and width of the patch to ensure the minimum resonant amplitudes of reflection coefficient, $|S_{11}|$ and isolation, $|S_{41}|$ at the center frequency of 6.5 GHz.

Slotlines in the ground plane introduce a slow wave structure, where the phase velocity of the propagating wave is smaller than the velocity of light. This makes a slotted ground plane become electrically longer. Since any additional phase shifter is not required in this work, this approach makes the proposed design different from the work reported in [26-27] and provides the main contribution in this work, especially in terms of bandwidth enhancements and size reduction.

The proposed 3 dB/45° patch coupler can operate at the center frequency of 6.5 GHz which is suitable for unlicensed used in 5.9 GHz to 7.1 GHz band that meets 5G spectrum requirements as supported by Federal Communications Commission (FCC) [2]. The proposed 3 dB/45° patch coupler is designed using Computer Simulation Technology (CST) Microwave Studio software as well as fabricated onto Rogers RO4003C board with substrate thickness, h of 0.813 mm and dielectric constant, ε_r of 3.38.

2. DESIGN ANALYSIS

Theoretically, power entering port 1 of the traditional 3 dB branch-line coupler is equally split between output ports, where the coupling factor is 3 dB with a 90° output phase difference. No power is coupled to the isolated port with all ports matched. The physical layout of the traditional 3 dB branch-line coupler with 90° output phase difference is depicted in Figure 1.

172

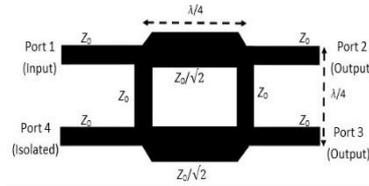Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 08, No. 02, June 2022.

Figure 1. Physical layout of traditional 3 dB branch-line structure with 90° nominal electrical length [36].

A block diagram of the traditional 3 dB branch-line coupler with an additional 45° phase shifter and the proposed 3 dB coupler without any additional phase shifter is illustrated in Figure 2 (a). The phase difference of the traditional 45° phase shifter can be determined by using the following expression [37]:

$$\text{Phase difference, } \Delta\phi = \frac{2\pi (L_m - L_r)}{\lambda_g} \tag{1}$$

where L_m, L_r and λ_g are the main length, reference length and guide wavelength, accordingly. The calculated L_m is 17.73 mm and has been optimized to obtain the desired 45° phase difference between the output ports (port 2 and port 3). The traditional 3 dB/90° branch-line coupler is combined with the traditional 45° phase shifter to develop 3 dB coupling value and 45° output phase difference. According to Equation (1), the electrical length of the main line will affect the value of the output phase difference. Figure 2 (b) shows that the 45° output phase difference for combination of the traditional 3 dB coupler with 45° phase shifter is calculated as $((\theta_{ab} + \theta_{be}) - (\theta_{ad} - \theta_{df}))$, when an input signal is excited at port 1. Meanwhile, the 45° output phase difference of the proposed 3 dB coupler without phase shifter is calculated as $((\theta_{a'b'}) - (\theta_{a'd'}))$, when an input signal is excited at port 1.

(a) (b)
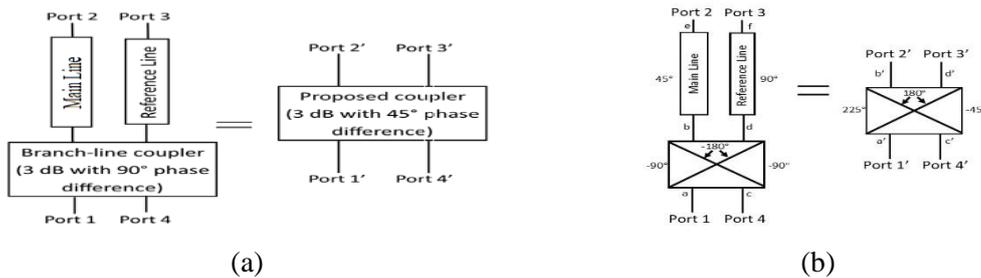
Figure 2. Traditional 3 dB coupler with 45° phase shifter and proposed equivalent: (a) Block diagram; (b) Signal paths.

The physical layout of the traditional 3 dB branch-line coupler with 45° phase shifter (Design A) is depicted in Figure 3. The parameters of Design A coupler are listed in Table 1. The objective of this work is to construct a single-layered 3 dB patch coupler with 45° output phase difference without using any phase shifter, which provides bandwidth enhancements and size reduction characteristics.
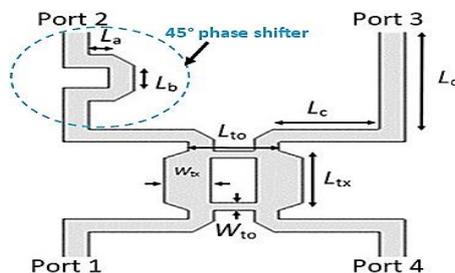
Figure 3. Physical layout of traditional 3 dB branch-line coupler with 45° phase shifter (Design A).

Table 1. Parameters of Design A coupler.

Parameters	Values (mm)
L_{to}	6.65
L_{tx}	7.45
L_a	1.46
L_b	2.91

L_c	7.3
L_d	14.59
W_{to}	1.02
W_{tx}	3.4

The proposed 3 dB/45° patch coupler (Design C) is implemented by modifying the 3 dB/90° patch coupler (Design B). The physical topology and photograph of Design B coupler under test are shown in Figure 4 (a) and Figure 4 (b), respectively.

(a)   (b)
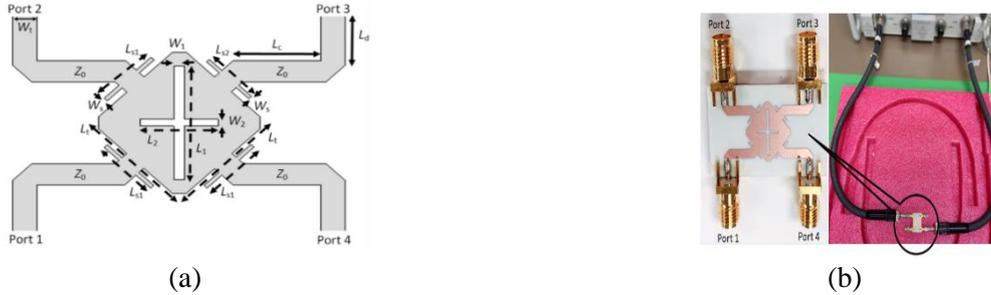
Figure 4. 3 dB/90° patch coupler with loaded stubs (Design B): (a) Physical topology; (b) Photograph under test.

Initially, the 3 dB/90° patch coupler is implemented by constructing a pair of cross slots on the square patch of the 3 dB/90° coupler (Design B). The cross slots are developed diagonally on the square patch to perturb the original patch resonator. Four feeding ports are placed in the middle of the square patch sides and matched to a characteristic impedance, Z_o which corresponds to 50 Ω. Each feeding port in Design B coupler is bent and extended to adapt with 0.5 λ_o inter-element spacing of the patch antenna array that can be integrated with Design B coupler to develop a switched-beam Butler matrix. The dimensions of length, L_t and width, W_t for the microstrip feeding ports can be calculated by substituting the dielectric method [31]. The calculated L_t equals a quarter-wavelength, whilst the calculated W_t is 1.88 mm. The dimensions of L_t and W_t are optimized to provide the optimum performance results. The initial widths of the cross slots, W_1 and W_2 are set to the available minimum fabrication limit of 0.2 mm. The capacitance and mismatch loss are reduced by chamfering each corner of the patch coupler. The initial lengths of the cross slots, L_1 and L_2 can be calculated using the Pythagorean theorem formula. Some parameters of the cross slots such as L_1, L_2, L_t, W_1 and W_2 are optimized to reroute the electric currents around the cross slots. According to the physical limitation of the square patch dimension, L_1 and L_2 are set to be less than $R = 9.98$ mm. Since it is not easy to convert the patch resonator into an analytical transmission line model, the physical layout is divided into four zones by the dash lines, as depicted in Figure 5. Each zone with different colours can be approximately corresponded to its equivalent transmission line model of the traditional branch-line coupler.

(a)   (b)
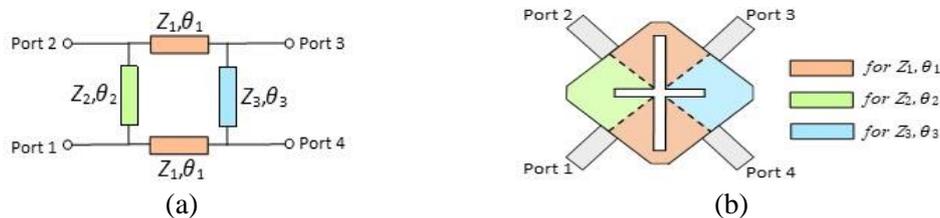
Figure 5. Schematic relationship: (a) Traditional branch-line hybrid coupler; (b) Cross-slotted patch coupler.

By applying the even-odd mode analyses [38]-[39], the transmission (T) and reflection coefficients (Γ) are obtained for each half circuit and the overall vector amplitudes of the signals emerging from the even-odd mode reflection and transmission coefficients could be determined as follows:

$$S_{11} = \frac{\Gamma_e + \Gamma_o}{2} \qquad (2)$$

$$S_{21} = \frac{T_e + T_o}{2} \qquad (3)$$

$$S_{31} = \frac{T_e - T_o}{2} \qquad (4)$$

$$S_{41} = \frac{\Gamma_e \text{-} \Gamma_o}{2} \tag{5}$$

The relationships between even-odd mode transmission coefficient equal to minus the odd-mode transmission coefficient are summarized in Equations (6)-(9). The phase difference between output ports is denoted by ψ.

$$S_{11} = S_{41} = 0 \tag{6}$$
$$S_{21} = T_e \tag{7}$$
$$S_{41} = \Gamma_e \tag{8}$$
$$\frac{S_{21}}{S_{31}} = \frac{\sqrt{(1\text{-}C^2)}}{C} \, e^{j\psi} \tag{9}$$

As ψ is defined between 0° and 90° to meet the arbitrary phase difference, the closed-form design equations can be further simplified as:

$$Z_1 = Z_0 \frac{\sqrt{(1\text{-}C^2)}}{C} \, \sin \psi \tag{10}$$

$$Z_2 = Z_0 \, \sin \psi \, \sqrt{\frac{(1\text{-}C^2)}{C^2 + \sin^2 \psi \, (1\text{-}C^2)}} \tag{11}$$

$$Z_3 = Z_2 \tag{12}$$
$$\theta_1 = 90° \tag{13}$$
$$\theta_2 = \cos^{-1} [-\sqrt{(1\text{-}C^2)} \cos \psi] \tag{14}$$
$$\theta_3 = 180° \text{-} \theta_2 \tag{15}$$

$$Z_2 = Z_0 \, \sin \psi \, \sqrt{\frac{(1\text{-}C^2)}{C^2 + \sin^2 \psi \, (1\text{-}C^2)}} \tag{16}$$

$$Z_3 = Z_2 \tag{17}$$
$$\theta_1 = 90° \tag{18}$$
$$\theta_2 = \cos^{-1} [-\sqrt{(1\text{-}C^2)} \cos \psi] \tag{19}$$
$$\theta_3 = 180° \text{-} \theta_2 \tag{20}$$

$\psi = 90°$ in both Design A and Design B couplers, whilst the calculated $Z_1 = 50 \, \Omega$, $Z_2 = Z_3 = 35.36 \, \Omega$ and $\theta_1 = \theta_2 = \theta_3 = 90°$. Nevertheless, it is difficult to realize the equivalent closed-form transmission line theory owing to the complicated field distribution on the cross-slot patch, as discussed in [40]. However, this can be achieved by adjusting the length and width of the inductive slots. Theoretically, the series inductance of the microstrip line is reduced, while the shunt capacitance is increased after converting the layout from the microstrip line into the patch structure. The cross slots' lengths, such as L_1 and L_2, are optimized to make the minimum frequencies of $|S_{11}|$ and $|S_{41}|$ be similar at 6.5 GHz.

Next, rectangular stubs are introduced at each feeding port to improve the bandwidth of S-parameter responses. The lengths of cross slots and stubs vary inversely with the frequency of propagation according to Equation (21) [31]. The square patch occupies an area of $0.2 \, \lambda_o \times 0.2 \, \lambda_o$. The final dimensions of Design B coupler are listed in Table 2. Design B coupler achieves a size reduction of 26.32% compared to the traditional 3 dB branch-line coupler.

$$\text{Frequency = Propagation velocity / wavelength} = v/(\lambda\sqrt{(\varepsilon\_r)}) \tag{21}$$

Table 2. Final dimensions of Design B coupler.

Parameters	Values (mm)	Parameters	Values (mm)
L_t	9.39	L_{s2}	4.53
L_1	9.85	W_t	1.84
L_2	5.92	W_1	0.80
L_c	6.65	W_2	0.60
L_d	4.00	W_s	0.30
L_{s1}	5.00		

After that, the design topology of Design B coupler in Figure 4 (a) has been modified to implement the proposed 3 dB patch coupler with 45° output phase difference without using 45° phase shifter (Design C). The design modification is realized by reducing the output phase difference from 90° to 45°. ψ is

"Bandwidth Enhancements and Size Reduction of 3 dB Patch Coupler with 45° Output Phase Difference for 5G Beamforming Networks",
N. S. M. Suhaimi, A. N. Uwaechia and N. M. Mahyuddin.

set to 45°, whilst the calculated $Z_1 = 35.36\ \Omega$, $Z_2 = Z_3 = 27.36\ \Omega$, $\theta_1 = 90°$, $\theta_2 = 120°$ and $\theta_3 = 60°$ using Equations (10)-(20). The design modification steps for the proposed 3 dB/45° patch coupler are realized as follows:

i. Dumbbell-shaped slots are introduced at the end of the cross slots, whilst lengths and widths of the patch, cross slots and rectangular stubs are optimized to reduce the output phase difference from 90° to 45°. The desired phase difference is controlled by selecting the suitable length and width of the dumbbell-shaped slots.

ii. The microstrip line is readjusted back to 50 Ω by optimizing the dimensions of chamfering corners while maintaining the performance of -10 dB reflection coefficient, $|S_{11}|$.

iii. Notches are placed along each length and width of the patch coupler by adopting the inset feeding technique to electrically tune the resonant amplitudes of reflection coefficient, $|S_{11}|$ and isolation, $|S_{41}|$ at the center frequency of 6.5 GHz as well as to improve the -3 dB $\pm$ 1 dB transmission coefficient, $|S_{21}|$.

iv. Rectangular slots are loaded at the ground plane (Design D) to improve the bandwidths of 45° $\pm$ 5° phase imbalance, -3 dB $\pm$ 1 dB transmission coefficient, $|S_{21}|$ and -3 dB $\pm$ 1 dB coupling coefficient, $|S_{31}|$.

The physical topologies of 3 dB/45° patch coupler without rectangular ground slot (Design C) and 3 dB/45° patch coupler with rectangular ground slots (Design D) are illustrated in Figure 6 (a) and Figure 6 (b), respectively. The photograph of the proposed Design D coupler with the rectangular ground slots under test is depicted in Figure 7. The final dimensions of Design C and Design D couplers are listed in Table 3. The proposed Design D coupler occupies an area of 0.22 $\lambda_0 \times$ 0.23 λ_0, which contributes to a size reduction of 45.72% compared to Design A coupler.

This proposed work is recommended for the beamforming network such as Butler matrix without using any additional phase shifter, which contributes to overall size reduction. The miniaturization factor (MF) for the proposed 3 dB/45° patch coupler with rectangular ground slots (Design D) can be calculated using the following equation:

$$MF\ (\%) = \frac{A_{traditional} - A_{proposed}}{A_{traditional}} \times 100 \qquad (22)$$
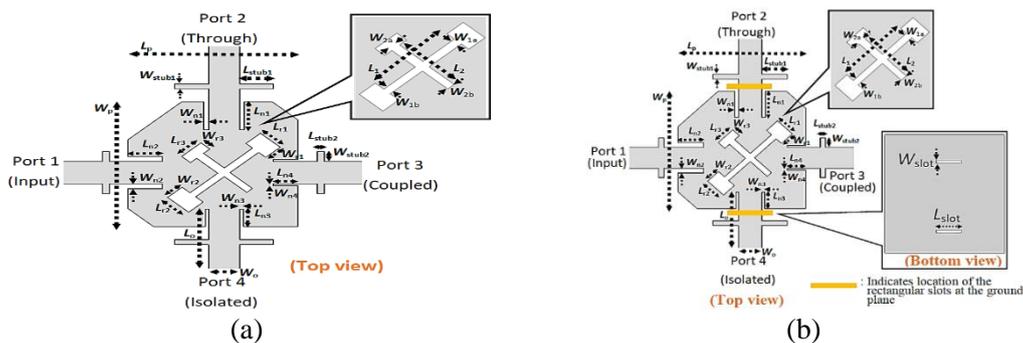
Figure 6. Physical layout of 3 dB patch couplers with 45° output phase difference: (a) Without rectangular ground slots (Design C); (b) With rectangular ground slots (proposed Design D).
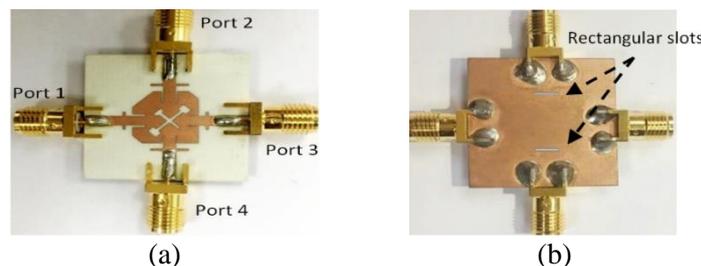
Figure 7. Photograph of fabricated Design D coupler with rectangular ground slots (proposed): (a) Top view; (b) Bottom view.

Table 3. Final dimensions of Design C and Design D couplers.

Parameters	Design C coupler	Design D coupler (proposed)
	Values (mm)	Values (mm)
L_o		7.50
L_p		10.12
L_1		5.31
L_2		4.14
L_{r1}		1.65
L_{r2}		1.78
L_{r3}		1.64
L_{stub1}		2.02
L_{stub2}		0.41
L_{n1}		2.36
L_{n2}		2.06
L_{n3}		1.48
L_{n4}		1.51
L_{slot}	-	3.98
W_o		1.84
W_p		10.50
W_{r1}		1.22
W_{r2}		1.20
W_{r3}		0.73
W_{1a}		0.51
W_{1b}		0.50
W_{2a}		0.59
W_{2b}		0.48
W_{stub1}		0.41
W_{stub2}		0.88
W_{n1}		0.29
W_{n2}		0.42
W_{n3}		0.20
W_{n4}		0.20
W_{slot}	-	0.38

3. RESULTS AND DISCUSSION

As observed in Figure 8, the -10 dB fractional bandwidths of the return loss, $|S_{11}|$ and isolation, $|S_{41}|$ for the integration of 3 dB branch-line coupler and 45° phase shifter (Design A) are 14.15% and 34.3% in simulation, whilst 10.77% and 23.07% in measurement, respectively. Meanwhile, the -3 dB ± 1 dB bandwidths of the respective transmission coefficient, $|S_{21}|$ and coupling, $|S_{31}|$ are 25.08% and 16.62% in simulation, whereas 23.07% and 16% in measurement. The 5° phase fluctuation bandwidths of the simulated and measured output phase differences are 12.77% and 13.23%, respectively. At the center frequency of 6.5 GHz, the simulated $|S_{11}|$ for Design A coupler is -13.59 dB, whereas the simulated $|S_{41}|$ is -15.3 dB. Both measured $|S_{11}|$ and $|S_{41}|$ are -18.23 dB and -14.38 dB, respectively. Meanwhile, the simulated $|S_{21}|$ is -3.31 dB, whereas the measured $|S_{21}|$ is -3.07 dB. Both simulated and measured $|S_{31}|$ are -3.71 dB and -3.93 dB, individually. The output phase difference is 45.45° in simulation, whereas 47.45° in measurement.

| (a) | (b) |
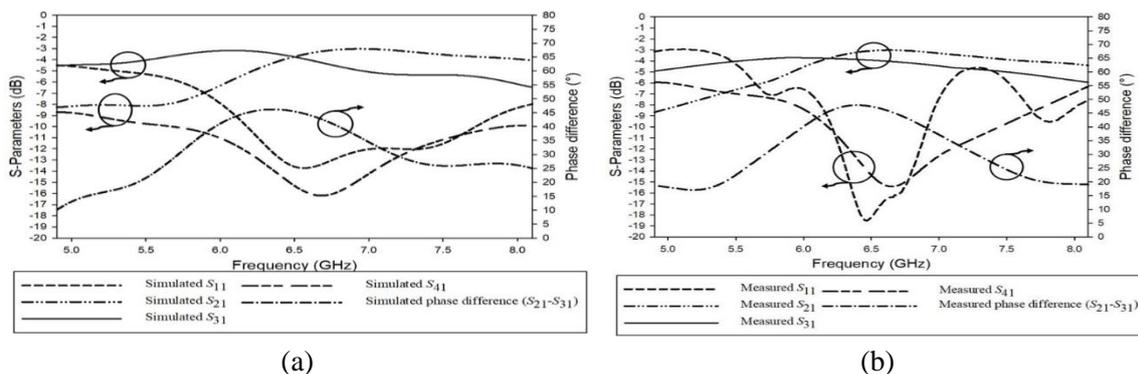
Figure 8. Integration of conventional 3 dB/90° branch-line coupler and 45° phase shifter (Design A): (a) Simulation results; (b) Measurement results.

"Bandwidth Enhancements and Size Reduction of 3 dB Patch Coupler with 45° Output Phase Difference for 5G Beamforming Networks", N. S. M. Suhaimi, A. N. Uwaechia and N. M. Mahyuddin.

Figure 9 shows the parametric analyses of variation cross slot's length, L_1 before loading the stubs and stub's length, L_{s1} after loading the stubs in Design B coupler. Figure 9 (a) shows that the values of the minimum peak for the matching response of $|S_{11}|$ are decreased and the operating frequency is shifted to the right by decreasing L_1. Although L_{s1} equals 3 mm, as shown in Figure 9 (b), provides the highest bandwidth of the $90° \pm 5°$ differential output phase, L_{s1} equals 5 mm is chosen in Design C, because it provides the highest fractional bandwidths of the simulated $|S_{11}|$, $|S_{21}|$, $|S_{31}|$ and $|S_{41}|$ compared to the other lengths of L_{s1}, as stated in Table 4.
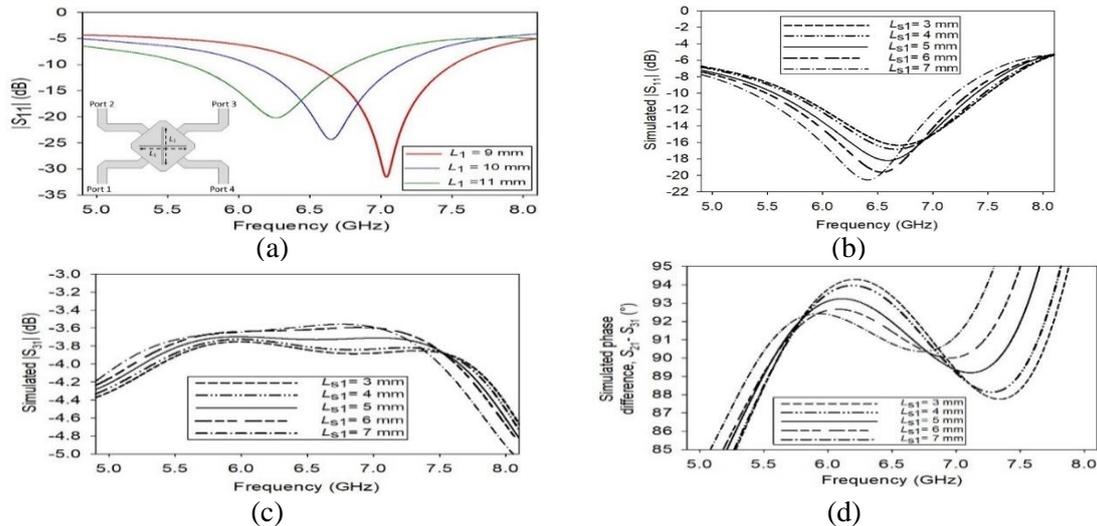
Figure 9. Parametric analyses of Design B coupler: (a) Simulated $|S_{11}|$ for the variation of cross slot's length, L_1; (b) Simulated $|S_{11}|$ for the variation of stub's length, L_{s1}; (c) Simulated $|S_{31}|$ for the variation of stub's length, L_{s1}; (d) Simulated phase difference, $|S_{21}| - |S_{31}|$ for the variation of stub's length, L_{s1}.

Table 4. Summarization of the proposed 3 dB/90° cross-slotted patch coupler with loaded rectangular stubs (Design C) performances based on the variations of stubs' lengths.

Variations of stub's lengths, L_{s1} (mm)	Bandwidth performance				
	$\|S_{11}\| \leq$ -10 dB	$|S_{21}| = $-3 $\pm$ 1 dB	$|S_{31}| = $-3 $\pm$ 1 dB	$\|S_{41}\| \leq$ -10 dB	Phase difference $= 90° \pm 5°$
3	24.15	18.77	35.85	38.92	39.54
4	27.23	18.15	36.15	39.69	38.92
5	28.00	19.85	36.46	40.92	37.84
6	27.84	18.77	36.00	40.76	35.08
7	28.00	19.08	36.15	39.38	33.85

The simulation and measurement results of S-parameters and output phase difference for the 3 dB/90° patch coupler with loaded stubs (Design B) are depicted in Figure 10. The -10 dB fractional bandwidths of the return loss, $|S_{11}|$ and isolation, $|S_{41}|$ for Design B coupler are 28% and 40.92% in simulation, whilst 30% and 36% in measurement, respectively. Meanwhile, the -3 dB $\pm$ 1 dB bandwidths of the respective transmission coefficient, $|S_{21}|$ and coupling, $|S_{31}|$ are 19.85% and 36.46% in simulation, whereas 18.46% and 34.77% in measurement. The 5° phase fluctuation bandwidths of the simulated and measured output phase differences are 37.84% and 38.62%, respectively. At 6.5 GHz, the simulated $|S_{11}|$ is -17.43 dB, whereas the simulated $|S_{41}|$ is -18.37 dB. Both measured $|S_{11}|$ and $|S_{41}|$ are -16.41 dB and -16.68 dB, respectively. Meanwhile, the simulated $|S_{21}|$ is -3.17 dB, whereas the measured $|S_{21}|$ is -3.32 dB. Both simulated and measured $|S_{31}|$ are -3.84 dB and -3.54 dB, individually. The output phase difference at 6.5 GHz is 90.82° in simulation, whereas 89.83° in measurement.
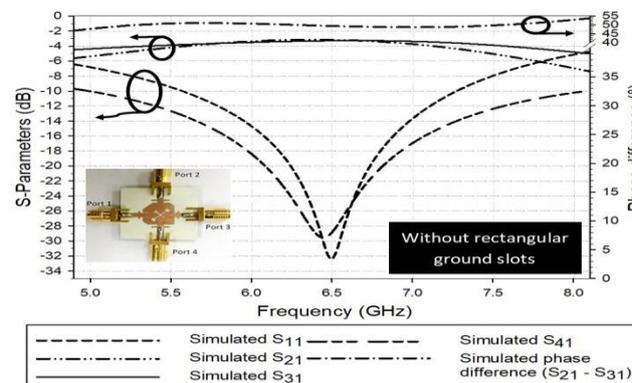
Figure 11 shows the simulation results of S-parameters and output phase difference for the 3 dB/45° patch coupler by applying step 1 to step 2 (Design C). The -10 dB fractional bandwidths of the simulated $|S_{11}|$ and $|S_{41}|$ for Design C coupler are 19.23% and 49.23%, respectively. The 1 dB amplitude fluctuation bandwidth for the simulated $|S_{21}|$ is 10.31%, whilst 35.07% for the simulated $|S_{31}|$. As observed, the

output phase difference of Design C coupler is capable to reduce from 90° to 45°. However, the bandwidth of $|S_{21}|$ is less than 18.5%. The bandwidth of 5° phase imbalance for the simulated output phase difference is 36.15%. At 6.5 GHz, the simulated $|S_{11}|$ is -18.76 dB, whereas the simulated $|S_{41}|$ is -19.73 dB. The simulated $|S_{21}|$ is -3.87 dB, whereas the simulated $|S_{31}|$ is -2.85 dB. Meanwhile, the simulated output phase difference at 6.5 GHz is 48.37° with a phase deviation of 1.63°.
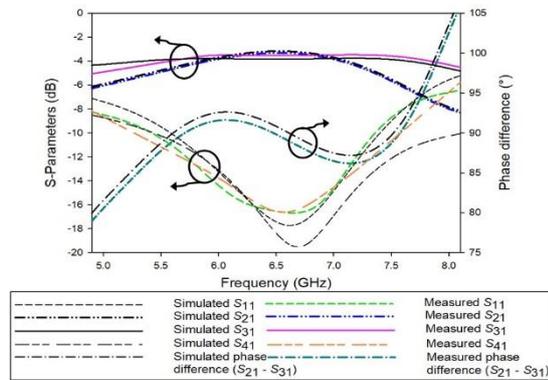
Figure 10. Simulation and measurement results of S-parameters and output phase difference for the 3dB/90° patch coupler with loaded stubs (Design B).
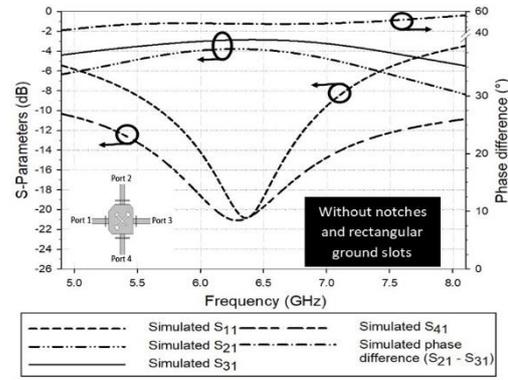
Figure 11. Simulation and measurement results of S-parameters and output phase difference for the 3dB/45° patch coupler by applying step 1 to step 2 (Design C).

Figure 12 shows the simulation results of S-parameters and output phase difference for the 3 dB/45° patch coupler by applying step 1 to step 3 (Design C). The -10 dB fractional bandwidths of the simulated $|S_{11}|$ and $|S_{41}|$ for Design C coupler are 26.02% and 47.07%, respectively. The 1 dB amplitude fluctuation bandwidth for the simulated $|S_{21}|$ is 23.69%, whilst 33.90% for the simulated $|S_{31}|$. As observed, the output phase difference of Design C coupler is capable to reduce from 90° to 45° without interfering with the performance of -3 dB coupling coefficient, maintaining the patch size by doing this modification. The bandwidth of 5° phase imbalance for the simulated output phase difference is 19.61%. At 6.5 GHz, the simulated $|S_{11}|$ is -32.37 dB, whereas the simulated $|S_{41}|$ is -28.86 dB. The simulated $|S_{21}|$ is -3.17 dB, whereas the simulated $|S_{31}|$ is -3.29 dB. Meanwhile, the simulated output phase difference at 6.5 GHz is 49.43° with a phase deviation of 4.43°. As observed, the presence of notches increases the -10 dB fractional bandwidth of the simulated $|S_{11}|$ and 1 dB amplitude fluctuation bandwidth for the simulated $|S_{21}|$. Moreover, the resonant amplitudes of reflection coefficient, $|S_{11}|$ and isolation, $|S_{41}|$ are electrically tuned at the center frequency of 6.5 GHz.

Figure 12. Simulation and measurement results of S-parameters and output phase difference for the 3dB/45° patch coupler by applying step 1 to step 3 (Design C).
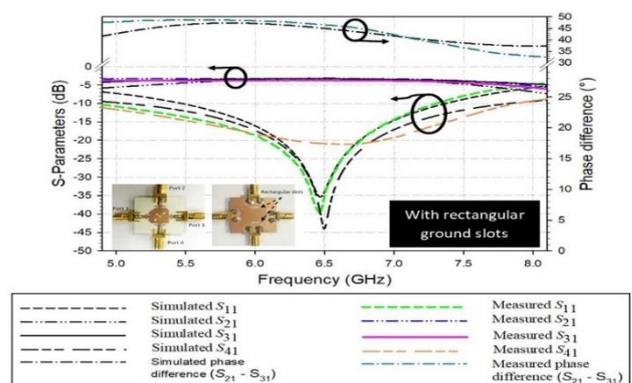
Figure 13. Simulation and measurement results of S-parameters and output phase difference for the proposed 3 dB/45° patch coupler by applying step1 to step 4 (Design D).

In order to improve the bandwidth performances of S-parameters and output phase difference from Design C coupler (without rectangular ground slots), rectangular ground slots are introduced in Design

"Bandwidth Enhancements and Size Reduction of 3 dB Patch Coupler with 45° Output Phase Difference for 5G Beamforming Networks", N. S. M. Suhaimi, A. N. Uwaechia and N. M. Mahyuddin.

D coupler. As observed, Figure 13 illustrates that the -10 dB fractional bandwidths of $|S_{11}|$ and $|S_{41}|$ for Design D coupler are 28.90% and 42.31% in simulation, whilst 37.08% and 46% in measurement. The respective -3 dB ± 1 dB bandwidths of $|S_{21}|$ and $|S_{31}|$ are 24.62% and 39.14% in simulation, whereas 42% and 37.08% in measurement. The output phase difference is 45° ± 5° between 4.9 GHz and 7.23 GHz which corresponds to 35.91% in simulation, whereas between 4.9 GHz and 7.18 GHz which corresponds to 35.08% in measurement. At 6.5 GHz, the simulated $|S_{11}|$ is -31.89 dB, whereas the simulated $|S_{41}|$ is -38.55 dB. Both measured $|S_{11}|$ and $|S_{41}|$ are -35 dB and -20.99 dB, respectively. The simulated $|S_{21}|$ is -3.17 dB, whereas the measured $|S_{21}|$ is -3.21 dB. Meanwhile, $|S_{31}|$ is -3.28 dB in simulation, whereas -3.71 dB in measurement.

The simulated output phase difference at 6.5 GHz is 45° without phase deviation, whereas the measured output phase difference is 46.62° with a phase deviation of ± 1.62°. The tolerances of measurement results are caused by the junction discontinuities as well as inaccuracy of measurement instrument and fabrication. The bandwidth enhancements of -3 dB ± 1 dB coupling coefficient, $|S_{31}|$ and 45° ± 5° output phase difference for Design D coupler are 22.52% and 23.14% compared to Design A coupler, respectively. Meanwhile, the bandwidth of the output phase difference in Design D coupler is improved by 16.3% owing to the presence of rectangular ground slots compared to Design C coupler.

4. COMPARATIVE STUDY

A performance comparisons of the 3 dB/45° patch coupler with rectangular ground slots (Design D) with other related works is summarized in Table 5.

Table 5. Performance comparison among the couplers (Design A, Design B, Design C and Design D) and other related works.

	Design A	Design B	Design C	Design D (proposed)	[31]	[32]	[33]
-10 dB bandwidth of $|S_{11}|$ (%)	14.15	28	26.02	28.90	21.40	22	10
1 dB amplitude imbalance bandwidth of $|S_{31}|$ (%)	16.62	36.46	33.90	39.14	24.50	42	30
5° phase imbalance bandwidth (%)	12.77	37.84	19.61	35.91	39.60	32	31
Electrical size ($\lambda_g \times \lambda_g$)	0.76×0.72	0.20×0.20	0.22×0.23	0.22×0.23	0.32×0.32	0.31×0.14	0.32×0.32

5. CONCLUSIONS

In this paper, a single-layered 3 dB/45° patch coupler with rectangular ground slots (Design D) is proposed. The inductive loading effect of cross slots in the proposed Design D coupler results in a significant size reduction of 45.72% compared to Design A coupler. The fractional bandwidths for -10 dB of $|S_{11}|$, -3 dB ± 1 dB of $|S_{31}|$ and 5° phase imbalance of output phase difference in Design D coupler are 28.90%, 39.14% and 35.91%, respectively. Moreover, the proposed Design D coupler provides bandwidth enhancements of 22.52% and 23.14% for -3 dB ± 1 dB coupling coefficient, $|S_{31}|$ and 45° ± 5° output phase difference compared to Design A coupler, respectively. The bandwidth of 45° ± 5° output phase difference in Design D coupler is increased by 16.3% owing to the presence of rectangular ground slots compared to Design C coupler. Therefore, the proposed 3 dB/45° patch coupler with rectangular ground slots (Design D) will be the attractive candidate in the future 5G beamforming networks such as Butler matrix without using additional phase shifters, which contributes to an overall size reduction while retaining reliable performance.

REFERENCES

[1]   I. A. Rumyancev and A. Korotkov, "Survey on Beamforming Techniques and Integrated Circuits for 5G

180

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 08, No. 02, June 2022.

Systems," Proc. of the IEEE International Conference on Electrical Engineering and Photonics (EExPolytech), pp. 76-80, St. Petersburg, Russia, 2019.

[2]   Federal Communications Commission, "Unlicensed Use of the 6 GHz Band, Notice of Proposed Rulemaking," Federal Registers, pp. 31997-32002, Washington, D.C, 2020.

[3]   H. Nachouane, A. Najid, A. Tribak et al., "Broadband 4×4 Butler Matrix Using Wideband 90° Hybrid Couplers and Crossovers for Beamforming Networks," Proc. of the International Conference on Multimedia Computing and Systems (ICMCS), pp. 1444-1448, Marrakec, Morocco, 2014.

[4]   D. N. A. Zaidel, S. K. A. Rahim and N. Seman, "4×4 Ultra-Wideband Butler Matrix for Switched Beam Array," Wireless Personal, vol. 82, no. 4, pp. 2471-2480, 2015.

[5]   S. F. Ausordin, S. K. Abdul Rahim, N. Seman, R. Dewan and B. Sa'ad, "A c3.3ompact 4×4 Butler Matrix on Double-layer Substrate," Microwave and Optical Technology Letters, vol. 56, no. 1, pp. 223-229, 2013.

[6]   R. Liang, Y. Zhang, W. Yan et al., "A Novel Design of Miniaturized Butler Matrix," Proc. of the 18th International Conference on Communication Technology (ICCT), pp. 560-563, Chongqing, China, 2018.

[7]   K. Han, W. Li and Y. Liu, "Flexible Phase Difference of 4×4 Butler Matrix without Phase-shifters and Crossovers," International Journal of Antennas and Propagation, pp. 1-7, 2019.

[8]   B. Li, Y. Chen, Z. Fu et al., "Substrate-guided Wave Optical True-time-delay Feeding Network for Phased-array Antenna Steering," Proc. of the International Society of Optics and Photonics (SPIE), pp. 256-265, San Jose, CA, United States, 2000.

[9]   F. Casini, R. V. Gatti, L. Marcaccioli et al., "A Novel Design Method for Blass Matrix Beamforming Networks," Proc. of the 37th European Microwave Conference (EuMA), pp. 1512-1514, Munich, Germany, 2007.

[10]   P. Chen, W. Hong, Z. Kuai and J. Xu, "A Double Layer Substrate Integrated Waveguide Blass Matrix for Beamforming Applications," IEEE Microwave and Wireless Components Letters, vol. 19, no. 6, pp. 374-376, 2009.

[11]   W. Y. Lim and K. K. Chan, "Generation of Multiple Simultaneous Beams with a Modified Blass Matrix," Proc. of the Asia Pacific Microwave Conference (APMC), pp. 1557-1560, Singapore, 2009.

[12]   T. Djerafi, N. J. G. Fonseca and K. Wu, "Broadband Substrate Integrated Waveguide 4×4 Nolen Matrix Based on Coupler Delay Compensation," IEEE Transactions on Microwave Theory and Techniques, vol. 59, no. 7, pp. 1740-1745, 2011.

[13]   F. E. Fakoukakis and G. A. Kyriacou, "Novel Nolen Matrix Based Beamforming Networks for Series-fed Low SLL Multibeam Antennas," Progress in Electromagnetics Research, vol. 51, pp. 33-64, 2013.

[14]   D. M. Pozar. Microwave Engineering, 4th Edn., New York: John Wiley & Sons, Inc., 2012.

[15]   C. Gai, Y. Jiao and Y. Zhao, "Compact Dual-band Branch-line Coupler with Dual Transmission Lines," IEEE Microwave and Wireless Components Letters, vol. 26, no. 5, pp. 325-327, 2016.

[16]   W. Feng, Y. Zhao, W. Che et al., "Dual-/Tri-Band Branch Line Couplers with High Power Division Isolation Using Coupled Lines," IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 65, no. 4, pp. 461-465, 2018.

[17]   L. Piazzon, P. Saad, P. Colantonio et al., "Branch-line Coupler Design Operating in Four Arbitrary Frequencies," IEEE Microwave and Wireless Components Letters, vol. 22, no. 2, pp. 67-69, 2012.

[18]   A. M. Zaidi, M. T. Beg, B. K. Kanaujia et al. "A Dual Band Branch Line Coupler with Wide Frequency Ratio," IEEE Access, vol. 7, pp. 25046-25052, 2019.

[19]   L. Xia, J. Li, B. A. Twumasi et al. "Planar Dual-band Branch-line Coupler with Large Frequency Ratio," IEEE Access, vol. 8, pp. 33188-33195, 2020.

[20]   C. Gai, Y. C Jiao and Y. L. Zhao, "Arbitrary Power Division Quadrature Branch-line Coupler with Harmonic Suppression," Microwave and Optical Technology Letters, vol. 60, no. 1, pp. 256-260, 2017.

[21]   H.-R. Ahn and M. M. Tentzeris, "Arbitrary Power-division Branch-line Hybrids for High-performance, Wideband and Selective Harmonic Suppressions from $2f_o$," IEEE Transactions on Microwave Theory and Techniques, vol. 67, no. 3, pp. 978-987, 2019.

[22]   H.-J. Yoon and B.-W. Min, "Two Section Wideband 90° Hybrid Coupler Using Parallel-coupled Three-line," IEEE Microwave and Wireless Components Letters, vol. 27, no. 6, pp. 548-550, 2017.

[23]   A. Bekasiewicz, "Miniaturized Dual-band Branch-line Coupler with Enhanced Bandwidth," Microwave and Optical Technology Letters, vol. 61, no. 6, pp. 1441–1444, 2019.

[24]   U. Dilshad, C. Chen, X. Chen et al., "Broadband Quadrature Hybrid for Image Rejection in Millimeter-wave Receivers," Proc. of 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST), pp. 975-978, Islamabad, Pakistan, 2019.

[25]   J.-G. Chi and Y.-J. Kim, "A Compact Wideband Millimeter-wave Quadrature Hybrid Coupler Using Artificial Transmission Lines on a Glass Substrate," IEEE Microwave and Wireless Components Letters, vol. 30, no. 11, pp. 1037-1040, 2020.

[26]   W. Yangsheng, H. Fuping and K. Dewu, "Design of Broadband Planar Magic-T Using 3-dB Branch-line Coupler and Phase Shifter," Proc. of the 3rd Asia-Pacific Conference on Antennas and Propagation, pp. 81-83, Harbin, China, 2014.

"Bandwidth Enhancements and Size Reduction of 3 dB Patch Coupler with 45° Output Phase Difference for 5G Beamforming Networks",
N. S. M. Suhaimi, A. N. Uwaechia and N. M. Mahyuddin.

[27]   D. Letavin, S. Shabunin and D. Trifonov, "Investigation of the Connection of Additional Stubs to the Phase Shifter Based on the Directional Coupler," Proc. of the 2019 Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT), pp. 364-367, Yekaterinburg, Russia, 2019.

[28]   S. S. Hesari and J. Bornemann, "Design of a SIW Variable Phase Shifter for Beam Steering Antenna Systems," Electronics, vol. 8, no. 9, pp. 1-14, 2019.

[29]   W. Zhang, Z. Shen and S. J. Xuk, "A Compact Wideband Phase Shifter Using Slotted Substrate Integrated Waveguide," IEEE Microwave and Wireless Components Letters, vol. 29, no. 12, pp. 767-770, 2019.

[30]   S. Liu and F. Xu, "Novel Substrate-Integrated Waveguide Phase Shifter and Its Application to Six-port Junction," IEEE Trans. on Microwave Theory and Techniques, vol. 67, no. 10, pp. 4167-4174, 2019.

[31]   A. Singh and M. K. Mandal, "Arbitrary Coupling Arbitrary Phase Couplers with Improved Bandwidth," IET Microwaves, Antennas & Propagation, vol. 13, no. 6, pp. 748-755, 2019.

[32]   Y. Wu, L. Jiao, Q. Xue et al., "A Universal Approach for Designing an Unequal Branch-line Coupler with Arbitrary Phase Differences and Input/Output Impedances," IEEE Transactions on Components, Packaging and Manufacturing Technology, vol. 7, no. 6, pp. 944-955, 2017.

[33]   Q. He, C. Qi, C. Liu et al., "A Compact Arbitrary Power Division Coupler with Nonstandard Phase Difference," Journal of Electromagnetic Waves and Applications, vol. 32, no. 3, pp. 293-305, 2017.

[34]   S. A. Babale et al., "Single Layered 4 × 4 Butler Matrix without Phase Shifters and Crossovers," IEEE Access, vol. 6, pp. 77289-77298, 2018.

[35]   Y. -L. Li, Q. S. Liu, S. Sun and S. S. Gao, "A Miniaturised Butler Matrix Based on Patch Hybrid Couplers with Cross Slots," Proc. of the IEEE Antennas and Propagation Society International Symposium (APSURSI), pp. 2145-2146, Orlando, FL, USA, 2013.

[36]   M. Moubadir, H. Aziz, N. A. Touhami and A. Mohamed, "A Miniaturized Branch-line Hybrid Coupler Microstrip for Long Term Evolution Applications," Procedia Manufacturing, vol. 22, pp. 491–497, 2018.

[37]   P. Upadhyay, V. Sharma and R. Sharma, "Design of Microstrip Patch Antenna Array for WLAN Application," Int. J. of Engineering and Innovative Technology (IJEIT), vol. 2, no. 1, pp. 295-297, 2012.

[38]   Y. S. Wong, S. Y. Zheng and W. S. Chan, "Quasi-arbitrary Phase-difference Hybrid Coupler," IEEE Transactions on Microwave Theory and Techniques, vol. 60, no. 6, pp. 1530-1539, 2012.

[39]   Y. Wu, J. Shen and Y. Liu, "Comments on "Quasi-arbitrary Phase-difference Hybrid Coupler"," IEEE Transactions on Microwave Theory and Techniques, vol. 61, no. 4, pp. 1725-1727, 2013.

[40]   S. Sun and L. Zhu, "Miniaturised Patch Hybrid Couplers Using Asymmetrically Loaded Cross Slots," IET Microwaves, Antennas & Propagation, vol. 4, no. 9, pp. 1427-1433, 2010.

ملخص البحث:

في هذا البحث، يتمّ اقتراح نظام ربطٍ أحادي الطّبقة (3 ديسيبل/45 درجة) يرمز اليه بالتّصميم D لشبكات تشكيل الشُّعاع مـن الجيـل الخـامس (5G) باستخدام بُنيةٍ مشقوقة عرْضياً فـي هيئـة شقوق علـى شكل كُرتيْن بينهمـا قضيب، وأرومـات وصْـلٍ محمّلـة، وأثـلام، وشقوق أساسية مستطيلة. ويمتلك نظام الـرّبط المقتـرح (التّصميم D) القدرة علـى إلغـاء الحاجـة الـى مُزيحـات طوْر إضافية بزاوية 45 درجة في شبكات تشكيل الشُّعاع مثل مصفوفة بَتْلَـر، الأمـر الـذي يمثِّـل الإسهام الرئيسي لهذا البحث، وبخاصّةٍ فيما يتعلّـق بتقليـل الحجـم وتحسـينات عَـرْض النّطـاق. ويـؤدي التّصـميم D المقتـرح الـى تحقيـق نطاقـاتٍ جزئيّـة قـدرها 28.9% و 39.14% و 35.91% لِـ (-10 ديسـيبل) بالنّسبة الـى |S11|، و (-3 ديسـيبل± 1 ديسـيبل) بالنسـبة الـى |S31|، و 5 درجـات مـن عـدم الإتّـزان بالنّسبة الـى فَـرْق الطّـوْر فـي المخـرج. وتبلـغ التّحسـينات فـي عَـرْض النّطـاق (-3 ديسـيبل±1ديسـيبل) بالنسـبة الـى معامـل الـرّبط   S31 و (45±5 درجـات) بالنّسـبة الـى فـرق الطّـوْر فـي المخـرج التـي يحقّقهـا النّظـام المقتـرح (التّصميم D) 22.52% و 23.14% مقارنـةً بالتصميم A، علـى الترتيب. ويتحقق تحسين عَـرْض النّطـاق لفرق الطّـوْر فـي المخـرج (45±5 درجـات) بحيـث يـزداد بنسـبة 16.3% بسـبب وجـود الشُّقوق الأساسـية المسـتطيلة فـي التّصـميم D المقتـرح مقارنـةً بنظـام الـرّبط للتّصـميم C. يبلـغ حجـم التّصـميم المقتـرح (التّصـميم D) (λ_g 0.22 x λ_g 0.23) ويعنـي ذلـك أنّ الحجـم الكهربـائي لنظـام الربط المقتـرح (التّصميم D) هو أقلّ بنسبة (45.72%) مقارنةً بنظام الرّبط ذي التّصميم A.

182

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 08, No. 02, June 2022.

NEUROMORPHIC COMPUTING BASED ON STOCHASTIC SPIKING RESERVOIR FOR HEARTBEAT CLASSIFICATION

Chia Yee Saw and Yan Chiew Wong

ABSTRACT

Heart disease is the leading cause of mortality worldwide. The precise heartbeat classification usually requires a higher number of extracted features and heartbeats of the same class may also behave differently in patients. This will lead to computation overhead and challenges in hardware implementation due to the large number of nodes utilized in reservoir computing (RC) networks. In this work, a reservoir computing-based stochastic spiking neural network (SSNN) has been proposed for heartbeat rhythm classification, enabling a patient adaptable and more efficient hardware implementation with low computation overhead caused by minimum extracted features. Only a single feature is employed in template matching to achieve patient adaptability with minimal computation overhead. The single feature, QRS complexes, was extracted and fed into the neural reservoir with 20 neurons in a cyclic topology for arrhythmia similarity calculation and classification. 43 recordings of Electrocardiogram (ECG) signals that included both normal and arrhythmic beats from MIT-BIH arrhythmia database obtained from Physio-Net were used in this work. The proposed stochastic spiking reservoir achieves a sensitivity of 99.6% and an accuracy of 96.91%, signifying that the system is accurate and efficient in classifying normal and abnormal arrhythmias.

1. INTRODUCTION

According to the Centers for Disease Control and Prevention (CDC), heart disease caused the deaths of approximately 659,000 people in the United States each year [1]. The United States spent $363 billion on heart disease per year between 2016 and 2017. This comprises the expenditures of medical services and medications, as well as lost productivity as a result of mortality.

ECG signals are a useful instrument that has been widely employed in numerous applications for the investigation of cardiac diseases [2]–[4]. ECGs use electrodes to monitor the electric activity generated by the flexion and contraction of heart muscles [5]. The ECG results show the physical activity of the heart, indicating whether the heart is healthy or has an abnormal rhythm. Abnormal rhythm is often known as arrhythmia, which is the most common cause of cardiac illness.

In recent years, various computer-assisted diagnostic approaches for classifying heartbeats for arrhythmia prediction have been proposed. To achieve automated real-time classification approaches, lower computational cost and easy adaptability to hardware are necessary [6]-[8]. Neuromorphic computing is a potential alternative to conventional von Neumann computers for specialized sensory-processing or classification applications. Neuromorphic systems imitate the biophysics of neuro-biological networks by imitating the information processing mechanism of biological neurons and synapses [9]. Through repeating this basic cortical columnar arrangement of neurons and synapses, a biological brain discovers a cognitive computing pattern that is highly energy-efficient. Neuromorphic architectures are known for their ability to perform complex machine-learning tasks with high connectivity and parallelism on a smaller footprint more than conventional von Neumann systems [10]–[13]. These characteristics contribute to the implementation of neuromorphic architectures in hardware development.

However, a high number of features in ECG recognition leads to excessive hardware calculation

---

C. Y. Saw and Y. C. Wong are with Faculty of Electronic and Computer Engineering (FKEKK), Universiti Teknikal Malaysia Melaka (UTeM), Malaysia. Email: ycwong@utem.edu.my

183

"Neuromorphic Computing Based on Stochastic Spiking Reservoir for Heartbeat Classification", C. Y. Saw and Y. C. Wong.

overhead that most of the previous research works have paid less consideration to. C. Venkatesan proposed a k-Nearest Neighbours-based (kNN) arrhythmic beat classification approach that classified normal and abnormal ECG signals utilizing 14 time domain and frequency domain heart rate variability (HRV) features with an accuracy of 97.5 % [14]. In another study, Ye et al. employed multiclass classification with Support Vector Machine (SVM) with 5 features comprising morphological, wavelet, RR interval, independent component analysis (ICA) and principal component analysis (PCA) to differentiate between normal and abnormal signals [15]. Vedavathi et al. used the kNN algorithm to classify abnormal beats based on three features: higher-order statistics (HOS), spectral characteristics and temporal features [16]. S. Savalia et al. used deep neural network approaches, such as multi-layer perceptron (MLP) and convolution neural network (CNN) with three characteristics: P wave, QRS complex wave and T wave for arrhythmia classification [17]. S. Nahak suggested a fusion-based SVM for arrhythmia classification, which achieved a three-class classification accuracy of 93.33% [18]. Thus, in this work, template matching technique has been proposed using minimum feature extraction from the ECG signal while still maintaining accuracy in order to reduce the computational burden of the system, achieving efficient hardware implementation.

In terms of adaptability to hardware, field programable gate array (FPGA) is a potential hardware solution for this application, where a large portion of the electrical functionality inside the device can be customized with low implementation cost [19]-[21]. Mariel et al. proposed a parallel continuous neural network (CoNN) using FPGA for arrhythmia detection [19]. The categorization algorithms in prior study utilized characteristics including amplitude, phase and signal shape. The identifier is implemented using a total of 50284 lookup tables (LUT) and 471 flip-flops (FF). An FPGA-based back propagation neural network (BPNN) implementation was proposed by Egila et al. to categorize ECG signals [20]. The developed module employs Discrete Wavelet Transform (DWT) for feature extraction to extract four features while using significantly less hardware. W. Matthias et al. presented multiple MLP-based ECG anomaly detection implemented in Zynq FPGA; however, it still requires a lot of hardware resources. For instance, the proposed 10-6-2 12-bit multilayer perceptron requires 1613 FF and 1963 LUTs [21]. Although BPNN and MLP require significantly lower hardware resources, there are still difficulties in dealing with inter-patient variability. However, it must be noted that the hardware implementation of neural network suffers from resource constraints because of the typically large number of nodes utilized in RC networks and the high chip area required by the activities engaged with each processing node [22]. The proposed stochastic computing (SC) approach and simple cycle reservoir (SCR) architecture implemented in SSNN significantly minimized the number of connections, resulting in a more efficient hardware implementation.

Inter-patient variability is a crucial concern in ECG data, since heartbeats of the same class can behave differently in patients. Template matching approach was implemented to overcome the concern of inter-patient variability. Most of the research in the literature focused on simple arrhythmia classification, which categorizes arrhythmia based on distinct features without patient adaptability feature. In this work, we propose a high-performance arrhythmia detection system that uses stochastic spiking RC to have a minimal processing overhead and minimize inter-patient variability concerns. In contrast to other research in the literature, a high performance for arrhythmia classification is achieved by utilizing only one feature, the QRS complex.

Section 1 presents an extensive assessment of recent studies on arrhythmia classification, while Section 2 presents literature on spiking reservoir learning machine and architecture. Section 3 presents the proposed ventricular heartbeat classifier, including Pan Tompkins beat detection algorithm, feature extraction and stochastic neuron design. Section 4 presents the performances of R-peak detection and arrhythmia classification. Hardware resource utilization and feature size of the proposed classifier are then compared with those of previous studies. The conclusion is made in Section 5.

2. SPIKING RESERVOIR LEARNING MACHINE

Spiking reservoir network is a class of brain-inspired recurrent algorithms known as RC networks, aiming to reduce computational complexity and cost of training machine-learning models by using random, fixed synaptic strengths. In this work, the RC network has been built using a three-layer spiking neural network (SNN) to compute spikes encoded, a dynamic system and a readout mechanism and with little effort, it can achieve machine-learning functionality.

2.1 Spiking Neural Network (SNN)

Spiking neural networks are the most recent generation of Artificial Neural Networks (ANNs), in which neuron models communicate by sequences of spikes to process large amounts of data [14]-[15]. The information is encoded in the rate and timing of the spikes' arrival.

Figure 1shows two connected biological neurons which communicate through sequences of spikes. The incoming signals from connected neurons accumulated by dendrites are spatially and temporally summed by the cell's body. The neuron will generate a spike if it receives enough input that exceeds the threshold. The action potential is transmitted along the axon to other neurons of the nervous system. If the threshold is not surpassed, this indicates that insufficient input was received; the inputs fade quickly and no action potential is produced.

The conventional artificial sigmoidal neuron is used to model the spiking neuron, which uses firing rate to relay neural information. The sigmoidal unit's operation is depicted in Figure 2. A weight variable describing the strength of the influence on the postsynaptic neuron is used to simulate the synapse between two neurons. Positive and negative weights are used to represent excitatory and inhibitory synapses, respectively. The potential of a sigmoidal neuron is calculated by adding together all of the weighted firing rates of its presynaptic neurons. The activation function is used to determine the neuron's output from this potential which has a sigmoid form.
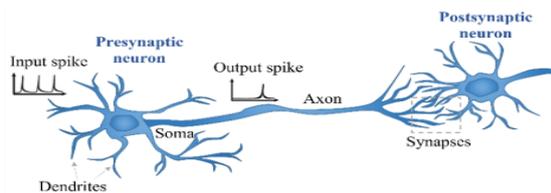
Figure 1. Biological neurons.
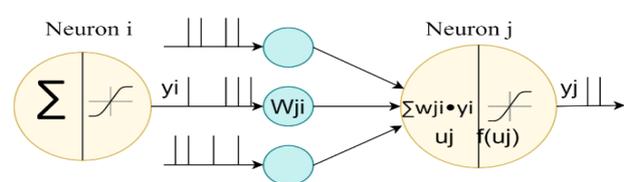
Figure 2. Basic operation of sigmoid neuron.

2.2 Reservoir Computing (RC)

RC is a computational architecture developed by recurrent neural network theory that maps input signals into higher-dimensional computational spaces utilizing the dynamics of a reservoir, which is a fixed, non-linear system. The reservoir is expected to be sophisticated enough to collect a significant number of input stream properties that can be utilized by the reservoir-to-output readout mapping [23].

Contrary to conventional Recurrent Neural Networks (RNNs) that need to adjust all connections to minimize the training error, the key characteristic of RC is that only the readout weights W_{out} are trained with a basic learning method such as linear regression, whereas the input weights W_{in} and recurrent connection weights W inside the reservoir are not trained [16], [18]. General RNN architecture and RC system are illustrated in Figure 3. This strategic design simplifies the complicated and time-consuming training procedure of traditional fully trained RNNs to a simple linear regression issue, which is greatly simplifying RNN implementation [19]-[20].
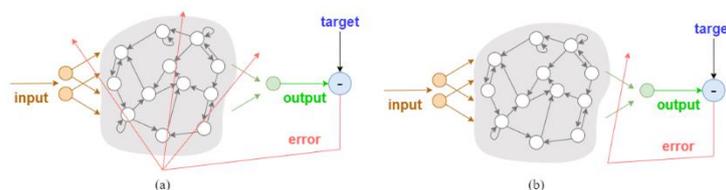
Figure 3. (a) General RNN architecture (b) RC system.

3. PROPOSED VENTRICULAR HEARTBEAT CLASSIFIER

Figure 4 depicts the procedures of the proposed technique. Signal denoising minimizes baseline drift and power-line interference that are typically emphasized in ECG signals. The peak detection detects the R-peak location and amplitude. The samples around each R-peak location are retrieved in a window by QRS segmentation. QRS complex is then employed in the spiking neural network for machine learning and classification. The correlation coefficient calculated from template matching was used to analyze the similarity between typical normal beats and the test beat. In this project, the well-known Pan

Tompkins beat detection algorithm is applied. This algorithm used band-pass filtering, signal differentiation, squaring, moving window integration and two sets of adaptive thresholds to filter and integrate signals for beat detection [24].

3.1 Signal Denoising

There are a variety of sounds and anomalies in polluted signals, but baseline drift and power-line interference [26] are two types of noise that are often highlighted in ECG signals. The impression of baseline wander (BW) occurs when a signal's base axis appears to "drift" up and down, causing the entire signal to deviate from its baseline. Baseline wander is a low-frequency distortion in the ECG caused by electrically charged electrodes, patient movement and respiration [27]. Power-line interference is another common cause of noise which must be removed from the signals, because it fully obliterates the low frequency P and T waves in ECG signals. These noises and artifacts can impede the extraction of useful information from the raw ECG signal, resulting in an incorrect diagnosis and having a significant impact on the performance of algorithms during classification [23], [28]-[29].

The initial step was to use band-pass filtering composed of a cascaded high-pass and low-pass filters with a passband of 5-15 Hz, to remove BW and 50 Hz power- line interference and reduce the T wave amplitude. Figure 5 shows the ECG signal pre-processing process.
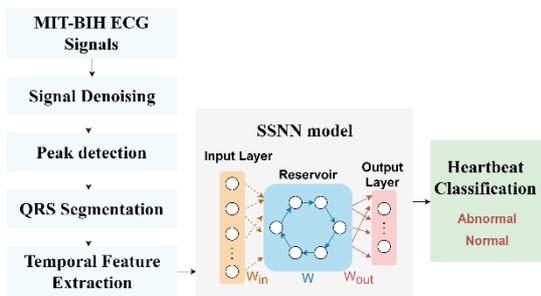
Figure 4. Schematic representation of ventricular heartbeat classifier [25].
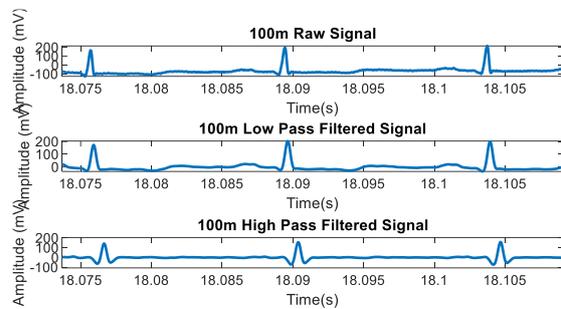
Figure 5. ECG signal denoising.

3.2 Peak Detection

The signal was differentiated after the band-pass filtering stage to highlight the severe slopes of the QRS complex. The squaring operation increases the slope of the derivative's frequency response curve to extract the R-peaks. The moving window integrator generates a signal including slope and QRS complex width information. The locations and amplitudes of R-peaks in the ECG signal were detected and stored in $R^{n \times m}$ matrix in a row-wise format, where $n = 2$ consists of location and amplitude of the R-peak and m is the number of detected R-peaks. Figure 6 shows an example of applying the Pan Tompkins algorithm for R-peak detection in the MIT-BIH arrhythmia database.
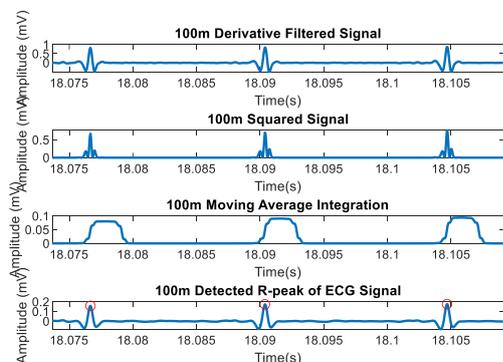
Figure 6. ECG signal peak detection.
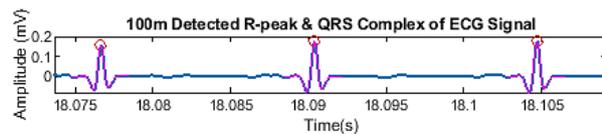
Figure 7. QRS segmented area in the ECG signal.

3.3 QRS Segmentation

The R-peak location contained in the $R^{n \times m}$ matrix is chosen as a reference point. After determining the R-peak location, 75 samples around each R-peak were extracted in the window and counted as a single

186

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 08, No. 02, June 2022.

heartbeat. The window only spans the duration of the QRS complex rather than the whole cardiac cycle. Figure 7 depicts the segmented region based on the QRS complex. The Q wave and S wave reflect the beginning and end point of the highlighted beat, whereas the R wave represents the peak area. The heart's ventricular depolarization is represented by the combination of Q, R and S waves [30].

3.4 Temporal Feature Extraction

The main interest of this research is to extract representative features with minimal calculation overhead and avoid inter-patient variability concerns. The simple and patient-adaptive technique template matching is utilized to alleviate the inter-patient variability difficulties outlined in Section 1. This approach is utilized to determine the similarity between the signal beat and the same class-specific template, then employed a normalized correlation coefficient defined as an equation to classify the data as shown in Figure 8.

$$\gamma_{xy}(k) = \frac{\sum_{n=0}^{N-1}[x(n)-\bar{x}][y(n-k)-\bar{y}]}{\sqrt{\sum_{n=0}^{N-1}[X(n)-\bar{x}]^2 \sum_{n=0}^{N-1}[y(n-k)-\bar{y}]^2}} \tag{1}$$

where γ_{xy} represents the correlation coefficient, N represents the number of template points, x(n) represents the template points, y(n) represents the signal points under investigation, x̄ represents the average of the template points and ȳ represents the average of the signal points and k represents the time index of the signal y(n) at which the template is placed. The correlation coefficient is in the range $-1 < \gamma_{xy} < +1$, where +1 means that the signal and template are perfectly matched.

Template matching approach is implemented in this work with a lower computation complexity while retaining a high compression ratio and a minimal reconstruction error. This study excludes from consideration those signals for the feature extraction phase that have contamination in their sections.
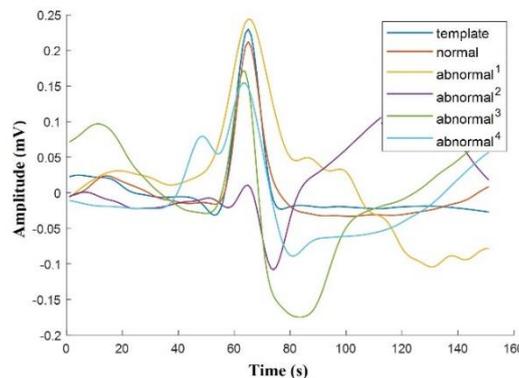
Figure 8. Template matching of segmented heartbeat.

The classification of normal and abnormal heartbeats is proposed by medical specialists in previous research [31]-[32]. Normal heartbeats have regular RR intervals, with the existence of a P-wave and a narrow QRS complex, whereas abnormal heartbeats have a shorter RR interval, no P-wave and a broader QRS complex.

3.5 Stochastic Spiking Neural Network

SSNNs are a recently proposed hardware solution based on a simple spiking neuron model capable of reproducing the probabilistic nature of synaptic transmissions, thus replicating the intrinsic stochastic behaviour of real biological neurons. The main distinction between SNNs and other neural networks is that SNNs explicitly model time. The concept is that neurons in the SNN do not communicate information at the end of each propagation cycle, but rather only when a membrane potential exceeds the threshold.

The retrieved QRS complex in the preceding section was employed as an input signal for the neural reservoir with 20 neurons. The stochastic neuron system contains three fundamental stages, as shown in Figure 9, binary to pulse conversion, SC and pulse to binary conversion. The input from the feature extraction process is converted into a binary number in the SSNN process by using analog-to-digital converter which transforms the analog input signal into the binary output signal. The binary value data

"Neuromorphic Computing Based on Stochastic Spiking Reservoir for Heartbeat Classification", C. Y. Saw and Y. C. Wong.

is first transformed into a pulse signal for stochastic computation and then the pulse signal is converted into its equivalent binary value again.
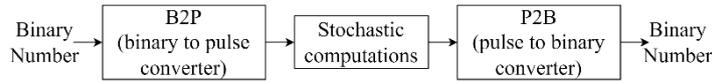
Figure 9. Stochastic computing system's fundamental stages.

3.5.1 Stochastic Computing (SC)

SC is a paradigm that counts the number of ones in a bitstream called a stochastic number, which is expressed as a fraction p=n/N, where n is the number of 1s in binary sequences and N is the length of binary sequences. For instance, the bitstream 01000010 contains two ones in an eight-bit stream, corresponding to a represented number of $p=P(X=1) = 2/8 = 0.25$. Figure 10 depicts the basic SC circuits, where the real number x is represented in unipolar format (UP) in the range [0, 1] and bipolar format (BP) in the range [-1, 1].

SC is utilized in this research by implementing digital gates in the SSNN. The proposed method reduces the amount of hardware required to perform arithmetic operations by employing probabilistic computing concepts. The multiplication in multiple-accumulate (MAC) operation could be easily realized by using a two-input XNOR gate for the bipolar coding format, as shown in Figure 10(b), followed by the scaled addition operation with a multiplexer (MUX), as shown in Figure 10(d).
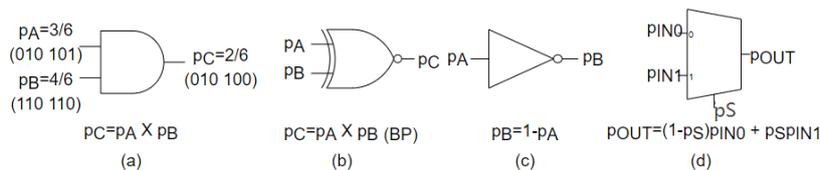
Figure 10. Basic circuits used in stochastic computation (a)AND (b)XNOR (c)NOT (d)MUX.

3.5.2 Binary-to-Pulse Converter (B2P)

The conversion from binary magnitudes (X) to pulsed stochastic signals (x) is performed by the B2P block shown in Figure 11. A random number generator (RNG) is used to create a pseudo-random binary number in each clock cycle, which is then compared to the n-bit input binary number with evaluation period T_{eval}. The comparator is actively high when $X > R_N$; otherwise it is maintained low. The comparator's output results in a bitstream $x = X/2^n$ with probability of getting 1.

3.5.3 Pulse-to-Binary Converter (P2B)

The P2B block is shown in Figure 11. Figure 12 is used to complete the process of converting the pulse signal x into its equivalent binary value X. It consists of two n-bit counters and an n-bit register. The first counter counts the number of high values (1s) given by the stochastic signal throughout the N_c clock cycle. The second counter is used to reset the first counter and load a new value for the register every N_c clock cycle. Therefore, the P2B block generates an n-bit number that changes per N_c cycle.

The block's output is a binary number that will remain stored for $N_c = 2n-1$ clock cycles until being updated with the next integration value. However, this conversion incorporates a statistical error. Thus, linear feedback shift register (LFSR) was implemented in this project and extended bit sequences were analyzed to obtain correct results. The probability of receiving an output corresponding to high-level values in a series of the random variable p(t) over Nc clock cycles is calculated using the binomial distribution equation:

$$Prob(X) = \binom{Nc}{X} p^X (1-p)^{Nc-X} \qquad (2)$$

3.6 Stochastic Neuron Design

The proposed stochastic neuron design consists of a few simple digital blocks: logic gates, a multiplexer, an LFSR, a register and a comparator. Figure 13 shows a single two-input sigmoid SC-based neuron with n=8 and evaluation time to $N_c = 2^8 - 1$ clock cycles.
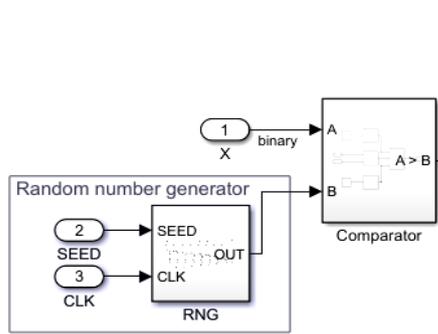
Figure 11. B2P converter block.
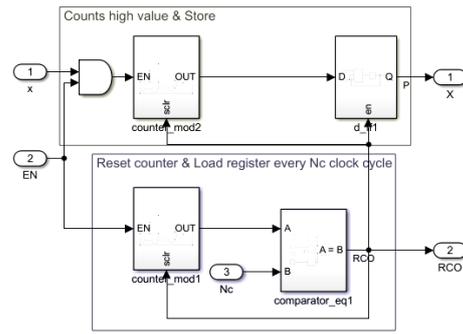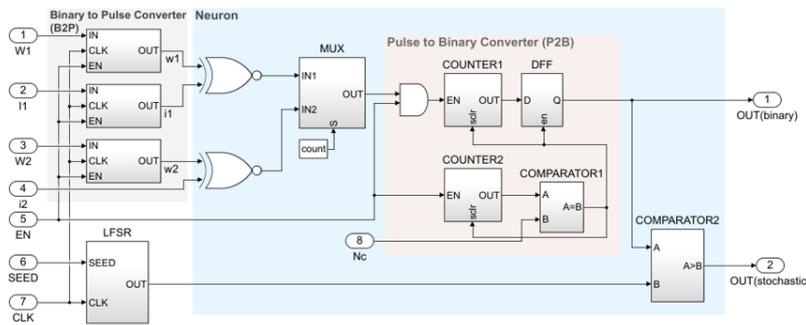

Figure 12. P2B converter block.

Figure 13. Two-input SC-based sigmoid neuron.

The input and weight binary values are first converted into pulsed signals using B2P converter before being processed by the stochastic circuit. The first input signal I_1 is a binary magnitude that is externally supplied to the system, but the second input signal i_2 is a stochastic bitstream that receives directly from another neuron. An XNOR gate and a multiplexer are used in the stochastic computing architecture to perform multiplication and addition operations. The resultant bitstream from the input weighting and addition is converted into a binary integer through P2B converter. The binary result is transformed back into a stochastic bitstream for further processing by another neuron.

20 stochastic neurons that make up the SSNN have been arranged based on SCR architecture as shown in Figure 14. This topology decreased the number of connections and simplified the automated network design, allowing for more efficient hardware implementation, when compared to previous research that employed random connections.
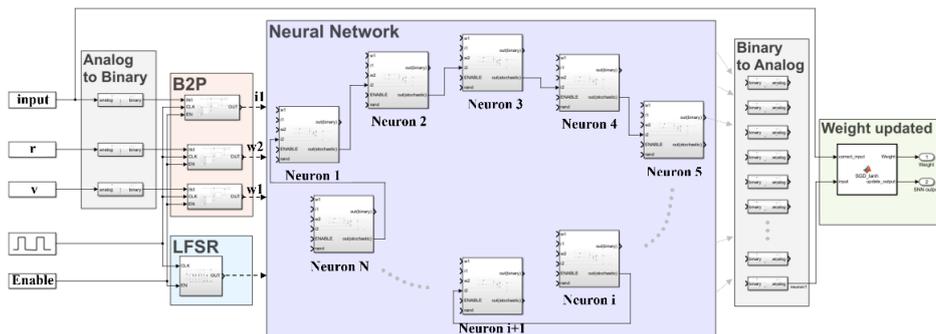
Figure 14. SSNN composed of 20 stochastic neurons in cyclic topology.

3.7 SNN Activity

The SNN's activity is depicted in   Figure 15. The input spikes entering the counter appear to contribute to an increase in the membrane over-voltage. During the evaluation process, the incoming pulses are added up per clock cycle T_{clk} in the counter, where the evaluation period is basically a definite number of clock cycles: $T_{eval} = N_c \cdot T_{clk}$. A random number generator is used to produce the value of the variable threshold V_k. The value of the estimated over-voltage corresponding to the preceding evaluation period is compared to the reference voltage per clock cycle. If membrane over-voltage S_k exceeds V_k, the

neuron emits a spike $X_k = 1$; otherwise it remains at a low level $X_k = 0$.

Figure 16 depicts QRS complex as the input signal associated with the output of various nodes in the reservoir. Each neuron state oscillates in response to the input stimuli and this information is utilised to calculate the network's prediction output. It can be observed that a good match occurs as the number of neurons trained increases.
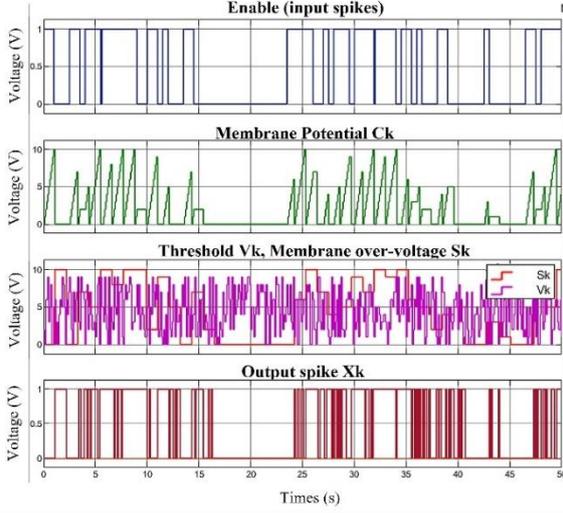
Figure 15. Temporal evolution of the different signals in the SNN model.
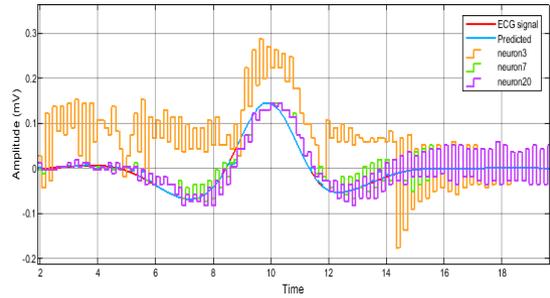
Figure 16. Three chosen neurons' output from the reservoir along with QRS complex input signal.

4. RESULTS, ANALYSIS AND DISCUSSION

The classification was carried out using the MIT-BIH arrhythmia database, which includes 43 recordings of ECG signals that included both normal and arrhythmic beats, each with a 30-min duration and a sampling rate of 360 Hz.

4.1 Performance of R-peak Detection

The Pan-Tompkins algorithm is used for R-peak detection and QRS segmentation. The performance of R-peak detection was evaluated using the following metrics: sensitivity (SEN), positive predictive value (PPV) and cumulative statistical index (CSI) and detection error rate (F_d) which are described in detail as follows:

$$SEN = \frac{TP}{TP+FN} \times 100 \tag{3}$$

$$PPV = \frac{TP}{TP+FP} \times 100 \tag{4}$$

$$CSI = \frac{1}{2}(SEN + PPV - FPR - FNR) \times 100 \tag{5}$$

$$F_d = \frac{FP+FN}{TP+FN+FP} \tag{6}$$

Table 1 reveals the overall performance of the R-peak detection on the MIT-BIH database, where out of a total of 97,963 beats, 205 and 227 beats were recognized as FP and FN, respectively. In addition, the proposed approach attained a minimum detection error of 0.44% with SEN, PPV and CSI at 99.8%, 99.8% and 99.6%, respectively. As a result, the Pan-Tompkins algorithm has an outstanding detection with minimal error detection, with high PPV indicating that a high proportion of R-peaks in the test beats are correctly detected, where high sensitivity indicates that the proposed approach is patient-adaptive and capable of detecting small outbreaks that behave differently in patients and larger CSI values indicate better overall detection performance.

Table 1. Evaluation results of R-peak detection.

	TP	FP	FN	SEN (%)	PPV (%)	CSI (%)	F_d (%)
Total	97963	205	227	99.8	99.8	99.6	0.44

True positive (TP) indicates a correctly detected R-peak, false negative (FN) indicates a missing R-peak and false positive (FP) indicates an incorrectly identified R-peak. False positive rate is FPR = FP/(FP+TP) and false negative rate is FNR = FN/(FN+TP). F_d was employed to calculate the detection error rate. Better detection performance is predicted by higher SEN, PPV and CSI values, while a lower F_d is expected. In Table 1, the TP, FP, FN, SEN, PPV, CSI and F_d values of each record are listed.

4.2 Arrhythmia Classification Performance

The proposed stochastic SNN architecture, which is made of 20 neurons and has a cyclic topology, was trained to complete a basic nonlinear prediction task. The evaluation duration has been set to $T_{eval} = (2^8 - 1) \cdot T_{clk}$ and the clock frequency is 1 kHz (T_{clk}=1ms). Figure 17 depicts the register transfer level (RTL) design. The network's performance was assessed for several patient datasets with QRS complex. Figure 18 depicts an example of one patient's ECG classification performance.
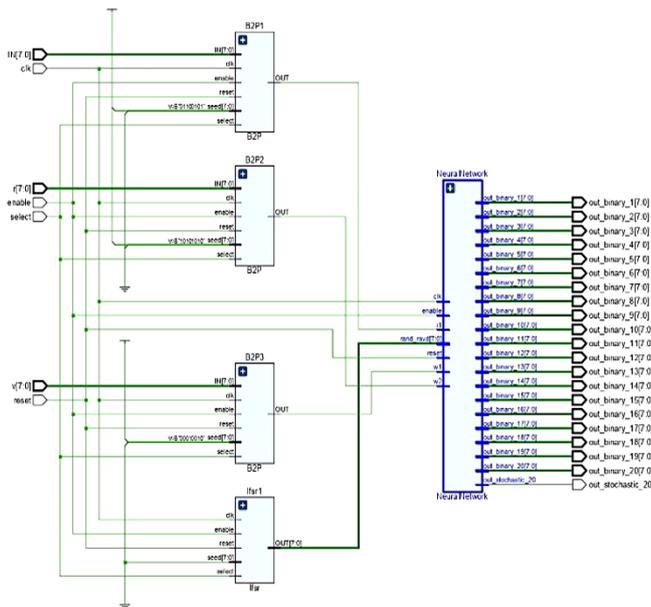
Figure 17. RTL of proposed SNN classifier.
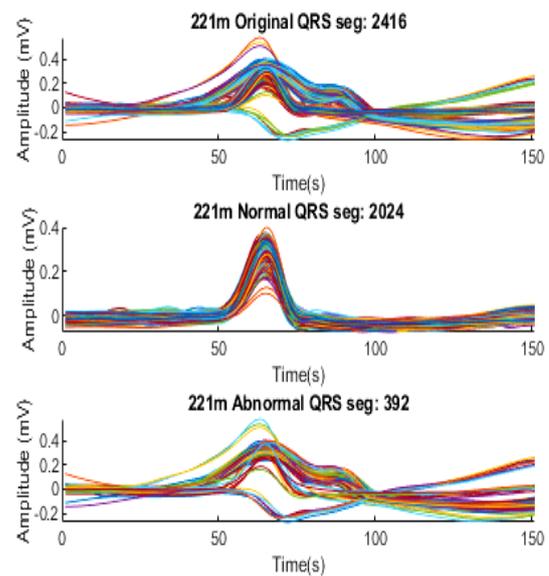
Figure 18. ECG classification performance.

Table 2 shows the confusion matrix for the performance of the proposed ventricular heartbeat classifier. The sensitivity, accuracy and specificity indicators are computed from the confusion matrix using Equations (3), (7) and (8) which are represents as follows.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \times 100 \tag{7}$$

$$Specificity = \frac{TN}{TN+FP} \times 100 \tag{8}$$

The template matching approach demonstrates that high performance was achieved with sensitivity, accuracy and specificity of 97.41%, 95.10% and 93.68%, respectively. The high accuracy indicates that the system is accurate and efficient in classifying normal and abnormal arrhythmias, while high specificity of abnormal diagnosis indicates the existence of disease with confidence.

Table 2. Confusion matrix of SNN classification.

	True Normal	True Abnormal
Prediction Normal	70085	3173
Prediction Abnormal	1862	27578

4.3 Comparison with Previous Study

Previous researchers have worked extensively on reducing the feature size of arrhythmia detection systems. However, they encountered computational burdens. Table 3 highlights the methodologies, feature size and classification performance achieved in the proposed methodology and related studies.

191

"Neuromorphic Computing Based on Stochastic Spiking Reservoir for Heartbeat Classification", C. Y. Saw and Y. C. Wong.

The proposed approach demonstrates that a great performance has been achieved by employing only a single feature as compared to previous research.

Table 3. Comparison with other studies.

The Authors	Method	Feature Size	Accuracy
Venkatesan et al. [14]	kNN	14	97.5%
Ye at al. [15]	SVM	5	86.4%
Vedavathi et al. [16]	kNN	4	98.40%
S. Savalia [17]	MLP; CNN	3	88.7%; 83.5%
S. Nahak et al. [18]	SVM	3	93.33%
Proposed Approach	**SNN**	**1**	**96.91%**

After verifying that the proposed method is effective in classifying heartbeats even with a single-feature size, the proposed system has been implemented in FPGA to demonstrate the hardware efficiency achieved by the minimum feature size used, thus reducing the computational burdens. Table 4 shows the hardware resources needed to implement the proposed SNN ventricular heartbeat classifier synthesized on a Xilinx Zynq-7000 FPGA. The proposed SNN classifier uses the least LUT and FF numbers to accomplish ventricular heartbeat classification, due to the hardware improvements in the MAC designs.

Table 4. Hardware resource utilization comparison.

Reference	[19]	[20]	[21]	**This Approach**
FPGA	Zynq 7000	Spartan 3AN	Zynq 7000	**Zynq 7000**
Classifier type	CoNN	BPNN	MLP	**SNN**
LUT	50284	4321	1963	**692 (0.65%)**
FF	471	3893	1613	**440 (0.83%)**

5. CONCLUSIONS

In this study, a simple and patient-adaptable heartbeat rhythm classification system based on SSNN was developed to obtain the highest performance with the minimum account load. The model only employs QRS complex as the only feature for classification. Moreover, the non-random structure of stochastic neurons connection in reservoir topology makes SSNN models easily implemented in hardware, also resulting in quick processing. The results demonstrate that the detection of QRS complex in MIT-BIH database using Pan-Tompkins and template matching achieves a sensitivity of 99.65%. Meanwhile, the proposed SSNN with correlation coefficient shows high effectiveness in classifying arrhythmia with sensitivity, specificity and accuracy of 99.6%, 98.93% and 96.91%, respectively. Simple processing and hardware resource saving of the architecture in this work provide a substantial contribution to the field of Medical Internet of Things (MIoT).

ACKNOWLEDGEMENTS

The authors acknowledge the technical and financial support by Universiti Teknikal Malaysia Melaka (UTeM) and the Ministry of Higher Education, Malaysia, under the research grant no. FRGS/1/2020/ICT02/UTEM/02/1.

REFERENCES

[1]   S. S. Virani et al., "Heart Disease and Stroke Statistics—2021 Update: A Report from the American Heart Association," Circulation, vol. 143, no. 8, pp. e254–e743, 2021.
[2]   K. Yasin, "Classification of PVC Beat in ECG Using Basic Temporal Features," Balkan Journal of Electrical and Computer Engineering, vol. 6, no. 2, pp. 78–82, 2018.
[3]   F. Zhou, L. Jin and J. Dong, "Premature Ventricular Contraction Detection Combining Deep Neural Networks and Rules Inference," Artificial Intelligence in Medicine, vol. 79, pp. 42–51, 2017.
[4]   X. Liu, H. Du, G. Wang, S. Zhou and H. Zhang, "Automatic Diagnosis of Premature Ventricular Contraction Based on Lyapunov Exponents and LVQ Neural Network," Computer Methods and Programs in Biomedicine, vol. 122, no. 1, pp. 47–55, 2015.
[5]   M. S. Memon, A. Lakhan, M. A. Mohammed, M. Qabulio, F. Al-Turjman and K. H. Abdulkareem,

"Machine Learning-data Mining Integrated Approach for Premature Ventricular Contraction Prediction," Neural Computing and Applications, vol. 33, no. 18, pp. 11703–11719, 2021.

[6] Y. C. Wong and Y. Q. Lee, "Design and Development of Deep Learning Convolutional Neural Network on an Field Programmable Gate Array," Journal of Telecommunication, Electronic and Computer Engineering, vol. 10, no. 4, pp. 25–29, 2018.

[7] C. Y. Saw, Y. C. Wong, S. L. Loh and H. Zhang, "On-chip Ultra Low Power Optical Wake-up Receiver for Wireless Sensor Nodes Targeting Structural Health Monitoring," Telkomnika, vol. 18, no. 5, pp. 2257–2264, 2020.

[8] Q. Mastoi, T. Ying Wah, R. Gopal Raj and A. Lakhan, "A Novel Cost-efficient Framework for Critical Heartbeat Task Scheduling Using the Internet of Medical Things in a Fog Cloud System," Sensors, vol. 20, no. 2, p. 441, 2020.

[9] F. Corradi et al., "ECG-based Heartbeat Classification in Neuromorphic Hardware," Proc. of the IEEE International Joint Conference on Neural Networks (IJCNN), pp. 1–8, Budapest, Hungary, 2019.

[10] A. F. Murray and A. V. W. Smith, "Asynchronous VLSI Neural Nnetworks Using Pulse-stream Arithmetic," IEEE J. Solid-State Circuits, vol. 23, no. 3, pp. 688–697, 1988.

[11] F. Blayo and P. Hurat, "A VLSI Systolic Array Dedicated to Hopfield Neural Network," Proc. of the VLSI for Artificial Intelligence, Part of the the Kluwer International Series in Engineering and Computer Science Book Series, vol. 68, pp. 255–264, Springer, 1989.

[12] P. Chi et al., "Prime: A Novel Processing-in-memory Architecture for Neural Network Computation in Reram-based Main Memory," ACM SIGARCH Comput. Archit. News, vol. 44, no. 3, pp. 27–39, 2016.

[13] K. Roy, A. Jaiswal and P. Panda, "Towards Spike-based Machine Intelligence with Neuromorphic Computing," Nature, vol. 575, no. 7784, pp. 607–617, 2019.

[14] C. Venkatesan, P. Karthigaikumar and R. Varatharajan, "A Novel LMS Algorithm for ECG Signal Preprocessing and KNN Classifier Based Abnormality Detection," Multimed. Tools Appl., vol. 77, no. 8, pp. 10365–10374, 2018.

[15] C. Ye, B. V. K. V. Kumar and M. T. Coimbra, "Combining General Multi-class and Specific Two-class Classifiers for Improved Customized ECG Heartbeat Classification," Proc. of the 21st IEEE International Conference on Pattern Recognition (ICPR2012), pp. 2428–2431, Tsukuba, Japan, 2012.

[16] V. G. Rangappa, S. Prasad and A. Agarwal, "Classification of Cardiac Arrhythmia Stages Using Hybrid Features Extraction with K-nearest Neighbour Classifier of ECG Signals," Learning, vol. 11, pp. 21–32, 2018.

[17] S. Savalia and V. Emamian, "Cardiac Arrhythmia Classification by Multi-layer Perceptron and Convolution Neural Networks," Bioengineering, vol. 5, no. 2, p. 35, 2018.

[18] S. Nahak and G. Saha, "A Fusion Based Classification of Normal, Arrhythmia and Congestive Heart Failure in ECG," Proc. of the IEEE National Conference on Communications (NCC), pp. 1–6, Kharagpur, India, 2020.

[19] M. Alfaro-Ponce, I. Chairez and R. Etienne-Cummings, "Automatic Detection of Electrocardiographic Arrhythmias by Parallel Continuous Neural Networks Implemented in FPGA," Neural Comput. Appl., vol. 31, no. 2, pp. 363–375, 2019.

[20] M. G. Egila, M. A. El-Moursy, A. E. El-Hennawy, H. A. El-Simary and A. Zaki, "FPGA-based Electrocardiography (ECG) Signal Analysis System Using Least-square Linear Phase Finite Impulse Response (FIR) Filter," J. Electr. Syst. Inf. Technol., vol. 3, no. 3, pp. 513–526, 2016.

[21] M. Wess, P. D. S. Manoj and A. Jantsch, "Neural Network Based ECG Anomaly Detection on FPGA and Trade-off Analysis," Proc. of the IEEE International Symposium on Circuits and Systems (ISCAS), pp. 1–4, Baltimore, MD, USA, 2017.

[22] A. Lakhan et al., "Hybrid Workload Enabled and Secure Healthcare Monitoring Sensing Framework in Distributed Fog-cloud Network," Electronics, vol. 10, no. 16, p. 1974, 2021.

[23] K.-L. Du and M. N. Swamy, "Recurrent Neural Networks," In book: Neural Networks and Statistical Learning, pp. 337–353, 2014.

[24] J. Pan and W. J. Tompkins, "Real-time QRS Detection Algorithm," IEEE Trans. Biomed. Eng., vol. BME-32, no. 3, pp. 230–236, 1985.

[25] Q. Mastoi, T. Y. Wah and R. Gopal Raj, "Reservoir Computing Based Echo State Networks for Ventricular Heart Beat Classification, Applied Sciences, vol. 9, no. 4, p. 702, 2019.

[26] G. M. Friesen, T. C. Jannett, M. A. Jadallah, S. L. Yates, S. R. Quint and H. T. Nagle, "A Comparison of the Noise Sensitivity of Nine QRS Detection Algorithms," IEEE Trans. Biomed. Eng., vol. 37, no. 1, pp. 85–98, 1990.

[27] R. Kher, "Signal Processing Techniques for Removing Noise from ECG Signals," Journal of Biomedical Engineering and Research, vol. 3, pp. 1–9, 2019.

[28] R. VanRullen, R. Guyonneau and S. J. Thorpe, "Spike Times Make Sense," Trends in Neurosciences, vol. 28, no. 1, pp. 1–4, 2005.

[29] M. L. A. Barceló, Methodologies for Hardware Implementation of Reservoir Computing Systems, Ph.D. Thesis, Physics Department, Electronics Engineering Group, Universitat de les Illes Balears, 2017.

"Neuromorphic Computing Based on Stochastic Spiking Reservoir for Heartbeat Classification", C. Y. Saw and Y. C. Wong.

[30] T. Y. Wah, R. Gopal Raj et al., "Reservoir Computing Based Echo State Networks for Ventricular Heart Beat Classification," Applied Sciences, vol. 9, no. 4, p. 702, 2019.

[31] H. Jaeger, "Tutorial on Training Recurrent Neural Networks, Covering BPPT, RTRL, EKF and the "Echo State Network" Approach," GMD-Forschungszentrum Informationstechnik Bonn, vol. 5, no. 01, 2002.

[32] F. Ponulak and A. Kasinski, "Introduction to Spiking Neural Networks: Information Processing, Learning and Applications," Acta Neurobiologiae Experimentalis, vol. 71, no. 4, pp. 409–433, 2011.

ملخص البحث:

تُعـدّ أمـراض القلْـب فـي طليعـة أسـباب الوفيـات علـى الصّـعيد العـالمي. ويتطلّـب التّصـنيف الـدّقيق لضـربات القلْـب عـدداً أكبر مـن السِّـمات المستخلَصـة، علمـاً بـأنّ ضـربات القلـب مـن الصِّـنف ذاتـه يُمكـن أن تتصـرّف علـى نحـوٍ مختلـفٍ مـن مـريض الـى آخـر. وهـذا مـن شـأنه أن يقـود الـى ارتفـاع فـي تكلفـة الحوسـبة، إضـافةً الـى تحـدّياتٍ تـرتبط بمعـدّات الحوسـبة المسـتخدمة، وذلـكَ بـالنّظر الـى العـدد الضّـخم مـن العُقَـد فـي شـبكات حوسـبة الخزّان (RC).

فـي هـذا البحـث، تـمّ اقتـراح شـبكةٍ عصـبية ناتئـة عشـوائية مبنيـةٍ علـى حوسـبة الخـزّان (SSNN) لتصـنيف إيقاعـات ضـربات القلـب، بحيـث تسـمح باسـتخدام معـدّات حوسـبةٍ ملائمـة للمرضـى بتكلفـةٍ حوسـبةٍ أقـلّ بسـبب اسـتخدام الحـدّ الأدنـى مـن عـدد السِّـمات المستخلصـة. وقـد تـمّ اسـتخدام سِـمَةٍ واحـدة فقـط لتحقيـق ملائمـة الشّـبكة للمرضـى وتقليـل تكلفـة الحوسـبة؛ تلـك السِّـمة هـي مركَّـب QRS، وقـد جـرى استخلاصـها وتغـذيتها الـى الخـزّان العصـبي بعشـرين عُصـبوناً علـى نحـوٍ دوْري مـن أجـل حسـاب وتصـنيف عـدم انتظـام ضـربات القلـب. وقـد تـمّ استخلاص 43 مـن تسـجيلات مخطّـط ضـربات القلـب المحتويـة علـى ضـربات قلـب طبيعيـة وأخـرى غيـر طبيعيـة مـن قاعـدة البيانـات الخاصـة بعـدم انتظـام ضـربات القلـب (MIT-BIH) التـي تـمّ الحصـول عليهـا مـن شـبكة (Physio-Net). وقـد حقـق الخـزّان المقتـرح حساسـيةٍ مقـدارها 99.6% ودقّـة مقـدارها 96.91%، الأمـر الـذي يؤشّـر الـى أنّ النّظـام المقتـرح دقيـق وفعّـال فـي تصـنيف حـالات عدم انتظام ضربات القلب الطبيعية وغير الطبيعية.

FULLY OPTIMIZED ULTRA WIDEBAND RF RECEIVER FRONT END

Rajesh Khatri and D. K. Mishra

ABSTRACT

This paper proposes a novel and fully optimized ultra-wideband RF receiver front end in UMC 180nm 1P6M CMOS process. The heterodyne architecture used in this work does not use the on-chip image reject mixer. The proposed receiver consists of a cascode inductively degenerated common source differential low noise amplifier and a folded Gilbert down-conversion mixer. The differential low-noise amplifier eliminates the use of active balun and improves the noise performance, while the folded architecture reduces the power dissipation of the receiver. The post-layout simulated result shows that the receiver has a voltage gain of 15.2 - 19.8dB, a noise figure of 4.8 - 8.9dB, a third-order input intercept point (IIP3) of -6.3 to -2.9dBm and consumes 31.5mW from a 1.8V supply. The receiver has a good reverse isolation S12 of -42 to -59dB due to cascode configuration and occupies an area of 2.55mm^2.

KEYWORDS

1. INTRODUCTION

Federal Communications Commission (FCC) has allocated a large frequency range of 3.1 to 10.6GHz (spectrum of 7.5GHz) for high-speed and short-distance communication. This ultra-wideband (UWB) IEEE 802.15.3a standard is used for wireless personal area network that transmits an extremely low signal power over a short distance at a high data rate (up to 480Mbps). Due to its ultra-wide bandwidth, people will popularly use it for sharing photos, music, videos, voice and data among laptops, PCs and mobiles connected in a network at home or office. There are two possible techniques to exploit the allocated spectrum. One is the multiband (MB) approach and the other is the Impulse-Radio (IR) approach [1]. In Multiband–UWB (MB-UWB), as specified in IEEE 802.15.3a standard, the entire bandwidth of 7.5GHz is divided into 5 band groups of 14 bands with a spacing of 528MHz, as shown in Figure 1, with OFDM (Orthogonal Frequency Division Multiplexing) modulation and frequency hopping scheme. The other possibility is the so-called Impulse-Radio-UWB (IR-UWB) based on transmission of very short pulses, with position or polarity modulation.
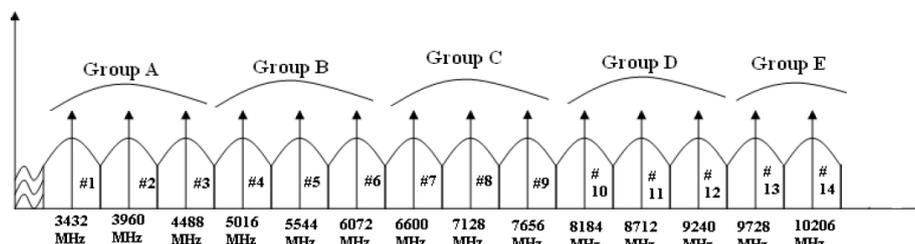
Figure 1. MB-UWB frequency plan.

UWB receiver front ends are challenging to researchers, as they require high receiver gain, high linearity, low signal-to-noise ratio (SNR) and minimum power consumption over a wide bandwidth. UWB is open to reception of undesired narrowband signals from WiFi and WiMAX systems under IEEE 802.11a/b/g as well as other UWB transmitters operating in the same range nearby. Due to narrowband jammers, nonlinearities present in the receiver can cause cross-modulation distortion, which further degrades the signal-to noise ratio. Hence, linearity (both IIP3 and IIP2) specification needs to include the distortion effects.

R. Khatri and D. K. Mishra are with Department of Electronics and Instrumentation Engineering, SGSITS, Indore (MP), India. Emails: rajeshkhatri1@rediffmail.com and dmishra@sgsits.ac.in

So far, several UWB receivers have been reported in literature [2]-[3]. In [2], a UWB direction conversion receiver for 3-5 GHz has been reported. In this research, the authors have achieved a receiver gain of 22dB, an NF of 17dB and a linearity (IIP3) of -6dBm while consuming 16mw from a supply of 2V in 0.18µm CMOS process. Although the receiver front-end has been fully integrated, it does not cover the full bandwidth of UWB from 3.1 to 10.6GHz and has a poor NF. The receiver front-end in [3] has been fabricated in 0.13µm Bi-CMOS process. The Bi-CMOS process has its advantages over the conventional CMOS process, but it is costly and consumes large power. Although the receiver implemented in this work shows a high gain of 52dB and a good IIP3 of -2.7 to -4.5dBm, it consumes a large power amounting to 88mw.

This paper describes designing and simulating of an optimized UWB receiver front end for narrowband and wideband jammers. Section 2 describes the receiver architecture. Specifications for the receiver are given in Section 3. Section 4 covers designing the LNA, mixer and BPF. The simulation results are discussed in Section 5, followed by a conclusion in Section 6.

2. RECEIVER ARCHITECTURE

Typical receiver architectures are direct-conversion receivers and heterodyne receivers. Direct conversion receivers are popular due to their simple architecture, low cost and high integrity. However, they have a problem of DC offset as well as sensitivity to narrowband jammers. Second-order distortion in base band can be another problem of direct-conversion receivers. Due to all these problems, there is a degradation of SNR in such receivers. Heterodyne receivers are less sensitive to second-order distortion and hence to SNR, but can possess various other problems like power consumption and image rejection. Image reject filter or image reject mixer is used for removing the image, but designing such mixer is a challenge. It requires an accurate quadrature local oscillator over a wide fractional bandwidth. In addition, it consumes significant power compared to traditional mixer. After considering the various pros and cons of direct-conversion and heterodyne receiver architectures, a new technique is proposed as an alternative to classical heterodyne receivers. In this proposed receiver architecture shown in Figure 2(a), the problem of image rejection is avoided by selecting an IF of 2.64GHz, so that all images fall below 2.64GHz. Therefore, an additional band pass filter with lower cut-off frequency of around 2.64GHz is used externally at input to remove all images as depicted in the frequency plan for the proposed receiver in Figure 2(b). Although wideband receivers normally use the wideband LO, the solid LO with wide IF has been selected to customize this heterodyne receiver, so that all the images fall below 2.64GHz as explained. Such UWB receivers find application in digital cameras and portable music players which frequently require data transfer of the order of a few gigabytes.

2.1 Design Methodology

The prime objective of this research is the optimization of performance parameters of UWB receiver front end. The performance parameters include receiver gain, noise figure, IIP3 and power dissipation. A comprehensive review of recently published works reveals that most researchers used the direct-conversion receiver due to its simplicity in addition to that it is, easy to design and requires less area due to on-chip components.
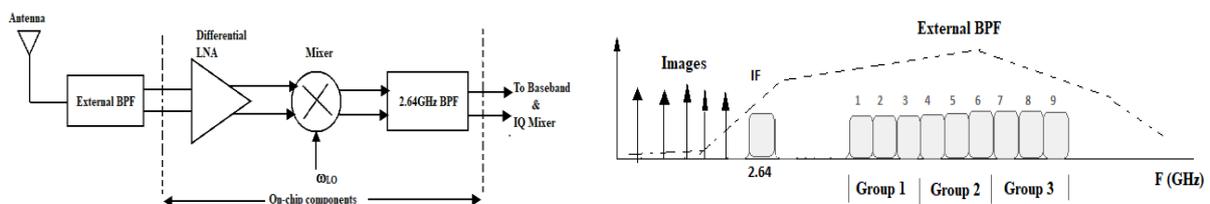

Figure 2(a). Proposed receiver architecture.　　Figure 2(b). Frequency plan for the proposed receiver.

The study also reveals that most authors avoided the use of on-chip inductors as they require more area and have a poor quality factor. The main research gap in these previously published works is that nearly all works were concentrated on reducing area while giving least attention to optimizing performance parameters, which is at most important, particularly at ultra-wide bandwidths. Some

papers have suggested the use of active inductors as an inductorless design to reduce the area. But, active inductors degrade the noise performance of receivers.

This paper tried to fill this research gap of optimizing performance parameters like gain, NF, IIP3 and power dissipation while giving less attention at area. The novelty of this work is that in this paper, a customized heterodyne receiver without image reject filter is designed without compromising on advantages of conventional heterodyne receivers as discussed above. The design methodology used in this work is based on the two well-known Frii's formulae [4]-[5] given by Equation 1 and Equation2 for cascaded blocks.

$$NF = NF_1 + \frac{NF_2 - 1}{G_1} + \frac{NF_3 - 1}{G_1 G_2} + \cdots \tag{1}$$

$$IIP3 = \frac{1}{\frac{1}{IIP3_1} + \frac{G_1}{IIP3_2} + \frac{G_1 G_2}{IIP3_3} + \cdots + \frac{G_1 \ldots G_n}{IIP3_n}} \tag{2}$$

Here, Equation 1 is for the overall NF of cascaded stages and Equation 2 is for the overall IIP3 of cascaded stages. From these equations, it is clear that the overall noise figure is dominated by the LNA, whereas the overall IIP3 is dominated by the mixer. Thus, there is a trade-off between noise figure and IIP3 which can be optimized by selecting their respective gain. Since this is an ultra wideband design, further optimization can be achieved by appropriately selecting the topology of LNA and mixer at the circuit level. In this design, differential inductively degenerated common source low-noise amplifier is selected to achieve input matching and low noise and to eliminate active balun. Similarly, a folded Gilbert mixer with load inductors is selected to achieve high IIP3 and low power.

3. RECEIVER SPECIFICATIONS

3.1 Receiver Linearity

As discussed in Section 2, UWB heterodyne receivers are sensitive to narrowband jammers, which in turn degrades the SNR of the receiver. In-band interferers include WiMAX and WiFi devices. In addition to these in-band interferers, a wideband jammer from another transmitting UWB system will produce cross-modulation distortion. UWB transmitter transmits an average power of -10.3dBm. If it is assumed that there is a noise figure of 8dB and an overall gain of 18dB for the LNA and mixer, then to overcome the cross-modulation distortion, the linearity; i.e., IIP3 of the receiver needs to be -8dBm to maintain the link margin of the receiver. Hence, the receiver was designed for an IIP3 of -8dBm.

3.2 Receiver Gain and Noise Figure

It is well known that there is a trade-off between gain and noise figure of the receiver. Generally, it is advantageous to provide a large gain in the receiver front-end, but this will increase the power consumption and degrade the linearity (IIP3). As the receiver consists of LNA and mixer, maximum gain will be provided by the LNA to ease the noise figure requirement on the mixer. Hence, the goal will be to keep front-end gain between 15 dB and 20dB.

4. LNA - MIXER DESIGN AND ANALYSIS

4.1 UWB LNA Design

Selecting the proper topology for the LNA, satisfying the specifications is a challenging task [6]-[7]. There are predominantly three topologies for wideband LNA design. These are: (1) Resistive shunt feedback amplifiers with input and output matching (2) Inductively degenerated common-source amplifier with LC input and output matching (3) Distributed amplifiers. In resistive feedback amplifiers, input matching is achieved by means of resistive termination. This gives a good matching, but the overall noise figure degrades. The power consumption of such circuits is also large. The distributed amplifiers are bulky, consume a large area and are power hungry.

Figure 3 depicts the schematic of the proposed ultra-wideband LNA. The differential configuration is used, which filters out the second-order harmonics and is suitable for a balanced mixer, which is also differential. This reduces the intermediate balun for converting the single ended output of the LNA

into a differential signal for the mixer. LNA uses an inductively degenerated common-source technique widely used in narrow-band designs with multisection reactive network, so that the overall input reactance is resonated over a wider bandwidth. As shown in Fig. (3), inductor Lp1 and capacitor Cp1 provide wide-band matching. Thus, a wideband input matching is achieved along with a good noise performance. To increase the flexibility in achieving perfect match, an inductor (Lg1) is placed in series with the gate of MOS transistor (M1) and a capacitor (Ct1) is also placed between the gate and the source of MOS transistor (M1). Here, the cascode configuration of MOS transistors M1 and M2 provides better reverse isolation (S12). It also improves the frequency response of the amplifier. MOS transistor M3 is used in source-follower configuration for buffering and to drive external load. Since differential configuration is used, the other half of the circuit is replicated. The main reason for using differential configuration for the LNA is to avoid the use of balun, which is bulky and may degrade the performance of the LNA.
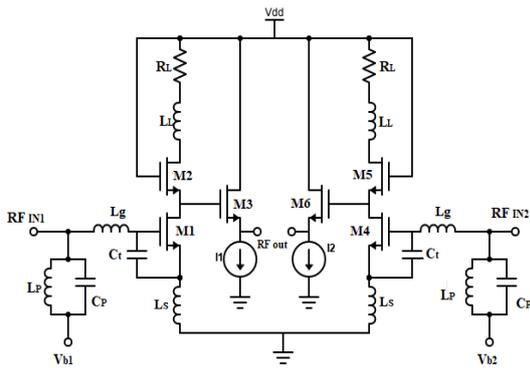
Figure 3. Proposed Ultra-Wideband LNA.
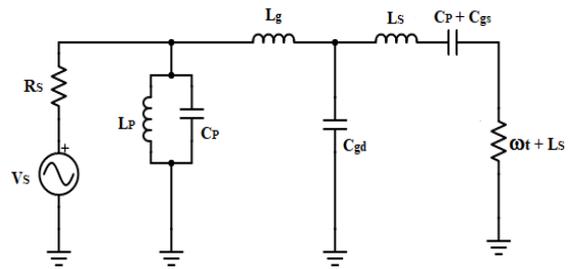
Figure 4. Section of LNA input network for impedance calculations.

4.1.1 Input-matching Analysis

Like in narrowband design, the reactive part of input impedance is resonated using two-section pass-band filter structure over the whole band from 3.1 to 8 GHz. Figure 4 depicts the section of LNA input network for input impedance calculations. As can be seen in Figure 4, the input impedance can easily be derived as given by Equation 3.

$$
\begin{aligned}
Z_{in}(s) &= \frac{1}{s(C_{gs} + C_p)} + s(L_s + L_g) + \omega_T L_s \\
&= \frac{s^2(L_s + L_g)(C_{gs} + C_p) + s\omega_T L_s(C_{gs} + C_p) + 1}{s(C_{gs} + C_p)}
\end{aligned}
\tag{3}
$$

where $\omega_T = \frac{g_m}{(C_{gs} + C_p)} = \frac{g_m}{C_t}$. The real part of Z_{in} is chosen to be equal to the source resistance; that is, $\omega_T L_s = R_s$. The gate-to-drain capacitance C_{gd} of MOS transistor M1 plays an important role in the resonance of the input circuit of the LNA. C_{gd} introduces one series resonance and one parallel resonance. The series resonance occurs between L_g and the parallel combination of L_s and C_{gd}. On the other hand, parallel resonance occurs between L_s and C_{gd}.

4.1.2 LNA Gain Analysis

To obtain the equation for the gain of the LNA, let us consider the transfer function of the filter section to be H(s). Hence, the input impedance is $\frac{R_s}{H(s)}$. Now, consider the current flowing into MOS transistor M1, which is, $v_{in}\left(\frac{R_s}{H(s)}\right)$. As we know, MOS transistor M1 acts as a current amplifier at high frequency with a current gain of $\beta(s) = \frac{g_m}{sC_t}$. Hence, the output current considering cascade stage is $\frac{v_{in}H(s)g_m}{(sC_t R_s)}$. Now, considering the load which is shunt peaking, the overall gain of the LNA can be obtained as given by Equation 4.

$$\frac{v_{out}}{v_{in}} = -\frac{g_m H(s)}{sC_t R_s} \frac{R_L\left(1 + \frac{sL_L}{R_L}\right)}{1 + sR_L C_{out} + s^2 L_L C_{out}} \tag{4}$$

where R_L is the load resistance, L_L is the load inductance and C_{out} is the total capacitance between the drain of MOS transistor M2 and ground.

4.1.3 LNA Noise Analysis

There are mainly two noise contributors in the proposed LNA. The first is the input network and the other is the noise of MOS transistor M1. Quality factor Q of the inductors in the input network decides the noise contribution from this network. The higher the quality factor Q for a given inductance value, the lower will be the noise. The noise cancellation techniques suggested in [8] can also be used for improving the noise performance of the LNA. Noise from MOS transistor M1 is due to drain current noise i_{nd} and gate-induced noise i_{ng}. Drain current noise is due to thermal agitation of carriers in the channel; while gate-induced noise is due to coupling of fluctuating channel charge into gate terminal. Noise due to both drain and gate is given by Equation 5 and Equation 6, respectively.

$$i_{nd} = 4kT\gamma g_{do} \tag{5}$$

$$i_{ng} = 4kT\delta \frac{\omega^2 C_{gs}^2}{5g_{do}} \tag{6}$$

where γ and δ are the noise parameters and g_{do} is the conductance for $V_{DS}=0$. There exists a correlation between drain current noise and gate noise. The correlation coefficient is represented by c and is $\approx j0.4$. By using classical noise optimization theory, the noise figure of the proposed UWB LNA can be derived as given by Equation 7.

$$F = 1 + \frac{P}{g_m R_s}\frac{\gamma}{\alpha} \tag{7}$$

where

$$P = \frac{p^2 \alpha^2 \chi^2 (1 - |c^2|)}{1 + 2|c|p\alpha\chi + p^2\alpha^2\chi^2} + \omega^2 C_t^2 R_s^2 (1 + 2|c|p\alpha\chi + p^2\alpha^2\chi^2)$$

where $P = {C_{gs}}/{C_t}$, $\chi = \sqrt{\delta/{(5\gamma)}}$, $\alpha = {g_m}/{g_{do}}$

Equation (5) gives the noise in the LNA due to MOS transistor M1, which is the main noise contributor. However, noise figure can be worse due to noise contributions from cascade MOS transistor M2 and output buffer MOS transistor M3. Considering noise and gain match along with the input match, the aspect ratio values of devices and values of input network components are given below in Table 1 and Table 2, respectively.

Table 1. Sizes of MOS devices.

No.	Device	Size
1.	MOS M1 and M4	280μm
2.	MOS M2 and M5	60μm
3.	MOS M3 and M6	100μm

Table 2. Values of components.

No.	Component	Values	No.	Component	Values
1.	Inductance L_p	2.5nH	5.	Capacitance C_t	100pF
2.	Capacitance C_p	340fF	6.	Load Inductance L_L	2.85nH
3.	Gate Inductance L_g	1.65nH	7.	Load Resistance R_L	110Ω
4.	Source Induct. L_s	720pH			

4.2 UWB Mixer Design

Down-conversion mixer is an important block in the UWB receiver. Like in the LNA, selecting a suitable down-conversion mixer for UWB range is a challenging task. Linearity of the UWB mixer decides the dynamic range of the receiver front end [9]-[10]. The linearity of the receiver is dominated by the mixer circuit. Many techniques have been suggested in literature for improving the linearity (IIP3) of the mixer. In [11], the third-order intermodulation (IM3) cancellation technique has been suggested for improving the IIP3 of mixer. Hence, UWB mixer should have good linearity, low noise and minimum power consumption. Various architectures for down-conversion mixers have been studied. Folded Gilbert mixer proposed in this paper satisfies the linearity, conversion gain and noise figure. It also consumes much less power due to its folded architecture.

Figure 5 depicts the schematic diagram of the proposed UWB down-conversion folded Gilbert mixer used in the receiver front end. The differential pair of NMOS transistors M1 and M2 forms the trans-conductance stage. PMOS transistors M3 through M6 are used in the LO stage, which is folded with respect to the trans-conductance stage. The output AC current from the trans-conductance stage should flow into the LO switches of PMOS transistors M3-M6. This can be achieved by using inductors L1 and L2 to provide high impedance. The folded architecture is preferred, as it significantly reduces power dissipation. NMOS Mb1 and Mb2 are used to provide biasing to the trans-conductance stage. The aspect ratio values of devices and values of components are provided in Table 3 and Table 4, respectively.

Table 3. Sizes of MOS transistors.

Sr. No.	Device	Size
1.	NMOS M1 and M2	50μm
2.	PMOS M3 – M6	100μm
3.	NMOS M0	240μm

Table 4. Values of components.

Sr. No.	Components	Values
1.	Inductors L1 and L2	5.5nH
2.	Bias Resistors R_b	440Ω

4.3 Band Pass Filter Design

An IF band pass filter has been designed to minimize the linearity requirements on the whole receiver chain and to eliminate the need for an external IF filter. Although on-chip BPF is not essential, it has been designed, as there may be multiple UWB transceivers operating in its vicinity, which may cause the problem of channel selection. There were two alternatives for filter selection; one is the active filter and the other is the passive filter. Active filters have traditional advantages of less area and excellent tuning over a wide range. But, they suffer from the main drawback of poor noise performance due to active components like MOS transistors. The noise performance is critical particularly at high frequency and wide bandwidth. The main objective of this research is to optimize the performance over cost (area). The passive filter has been selected. A traditional third-order Chebyshev band pass filter as depicted in Figure 6 has been designed.
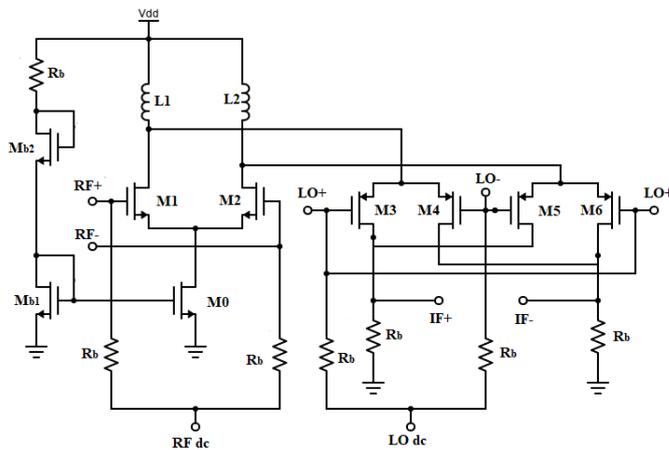
Figure 5. Proposed UWB folded Gilbert mixer.
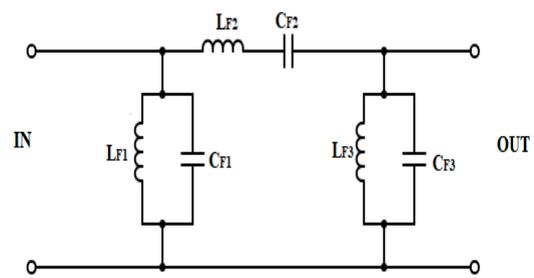
Figure 6. A typical LC band pass filter.

The filter with an IF center frequency of 2.64 GHz and a bandwidth of 528MHz was designed. This band pass filter improves the gain compression and intermodulation distortion of the IF downconverter. Table 5 gives the component values of the band pass filter.

Table 5. Values of components.

Sr. No	Components	Values
1.	L_{F1} and L_{F3}	520pH
2.	L_{F2}	2.5nH
3.	C_{F1} and C_{F3}	340fF
4.	C_{F2}	720pF

5. RESULTS AND DISCUSSION

The UWB receiver which consists of an LNA, a mixer and an IF band pass filter, is designed and simulated in 1P6M UMC 180nm CMOS technology. The complete schematic of the proposed UWB receiver is shown in Figure 7 and its layout in 1P6M UMC 180nm CMOS process is shown in Figure 8. The receiver occupies an active area of 2.55mm^2. As the proposed design contains a large number of on-chip inductors instead of active inductors, all of its performance parameters are optimized instead of area, which was the main objective of this research. In this section, schematic and post-layout simulation results of the LNA and the mixer are presented and discussed first, followed by receiver results. The heterodyne receiver is formed by directly connecting the output of the LNA to the mixer stage without any inter-stage matching for maximum voltage gain.
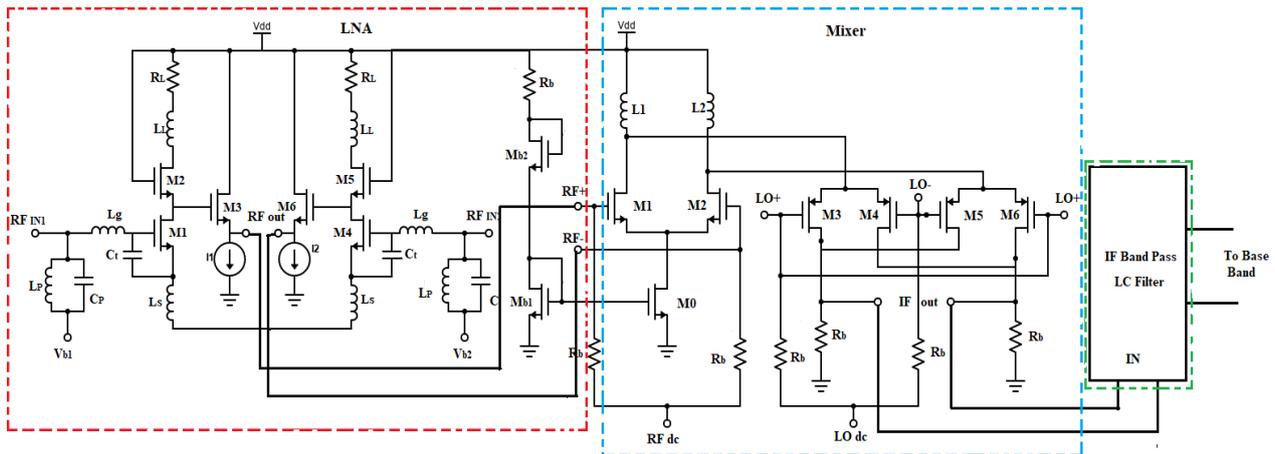
Figure 7. Complete schematic of the proposed UWB receiver front end.
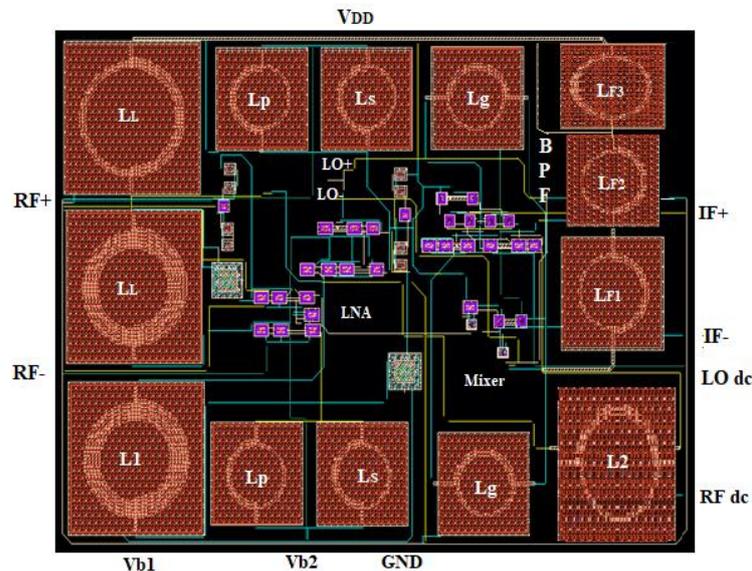
Figure 8. Layout of the proposed UWB receiver front-end.

5.1 LNA and Mixer Results

The post-layout simulated result of the UWB LNA gives excellent input matching (S11) of lower than -10dB over the entire range of 3.1 to 8GHz, whereas the gain (S21) of the LNA is from a minimum of 12.6dB to a maximum of 19.5dB, as shown in Figure 9. The fall of gain by 3dB is compared to schematic results. The fall in gain is due to parasitics generated while extracting the of layout. The LNA also shows a good noise figure of 5.1dB to 7.32dB over the entire range of interest, as shown in Figure 10. The noise figure also degrades by ±1dB due to parasitics of layout.

The schematic and post-layout simulation results depicted in Figure 11 show excellent reverse isolation (S12) of -66.8dB and -58.1dB, respectively, due to the cascode architecture used, which is

primarily used for improving isolation. The output matching (S22) as depicted in Figure 11 is also below -10dB in both schematic and post-layout simulation results. The LNA consumes 10.5mA at a supply voltage of 1.8V.
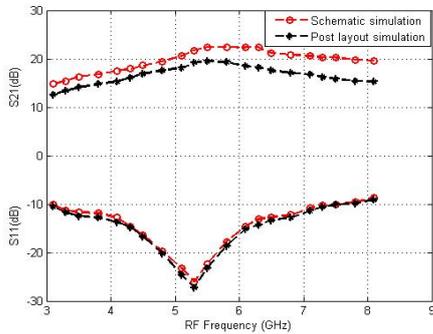
Figure 9. Simulated S11 and S21 of the LNA.
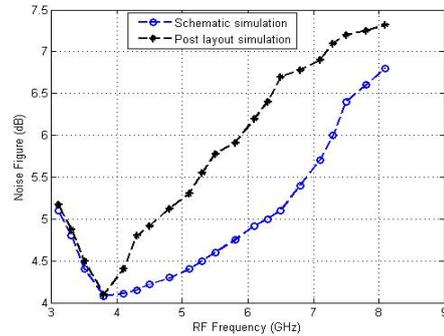
Figure 10. Simulated noise figure of the LNA.
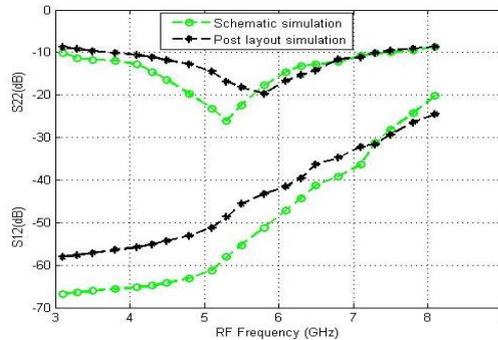
Figure 11. Simulated S12 and S22 of LNA.

Figure 12 shows the conversion gain (CG) and noise figure (NF) plot of the folded Gilbert mixer. The mixer has a schematic CG of 11.48dB to 14.68dB and a post-layout simulation CG of 10.78dB to
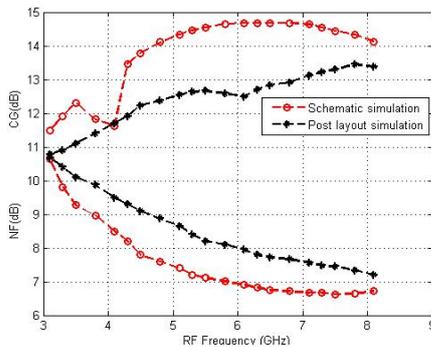
Figure 12. Simulated CG and NF of the mixer.
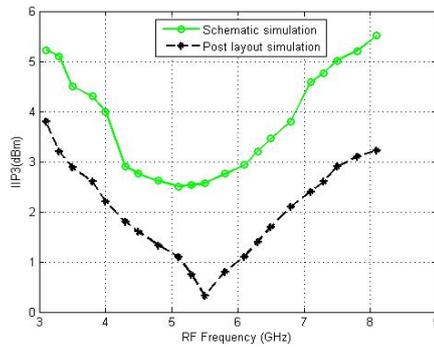
Figure 13. Simulated IIP3 of the mixer.
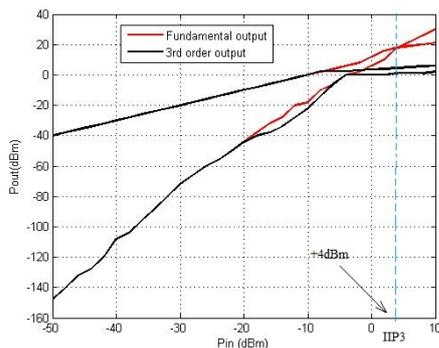
Figure 14. Simulation for IIP3 of the mixer at 4GHz.
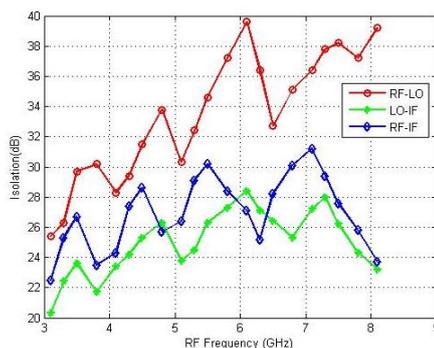
Figure 15. Mixer isolation *vs* RF frequency.

13.45dB over the entire band of 3.1 GHz to 8.1GHz. The mixer shows an NF of 6.8dB to 10.7dB for schematic simulation and an NF of 7.2dB to 10.6dB for post-layout simulation over the entire band of interest, which is acceptable when used with LNA in the receiver. Figure 13 depicts the IIP3 which is a measure of linearity of the mixer as a function of RF frequency. It shows excellent linearity of +2.5dBm to +5.5dBm in schematic simulation and slightly degraded linearity of +0.32 to +3.8dB in post-layout simulation. Figure 14 depicts the main signal power and third-order intermodulation power as function of the RF input power. The two-tone test is performed to calculate the IIP3 at a particular RF frequency. The two signals are fed to the RF input port, one at 4GHz and the other at 4.001GHz. The LO signal has a frequency of 3.9505GHz and a power level of -5dBm. The proposed folded Gilbert mixer exhibited an input third-order intercept point (IIP3) of +4dBm at an RF frequency of 4GHz. This result has been confirmed in Figure 13 as well. The port-to-port isolations of the proposed mixer were simulated and are presented in Figure 15. The port-to-port isolation represents the amount of leakage between the mixer ports. It is found that the mixer has an RF-LO isolation of better than 25dB, an LO-IF isolation of better than 20dB and an RF-IF isolation of better than 22dB. Due to folded architecture used in the mixer, it consumes only 7mA from a 1.8V supply.

5.2 Receiver Results

The UWB heterodyne receiver is implemented with the LNA, mixer and IF band pass filter as discussed in Section 2 with external LO signal. Due to differential architecture of the LNA, its output is directly connected to the mixer. The schematic simulation and post-layout simulation results of the complete receiver are presented. The receiver when simulated shows that the gain varies from 18.2dB in Band 1 (3.1GHz) to 22.8dB in Band 5 (5.5GHz) and is 19.6dB in Band 8 (7.1GHz), as shown in Figure 16. Similarly, the post-layout simulation results show that the gain varies from 15.4dB in Band 1 (3.1GHz) to 19.8dB in Band 5 (5.5GHz) and is 18.1dB in Band 8 (7.1GHz). The drop in the gain in the post-layout simulation is less than 1% to 2% and is well within the targeted specifications. This variation in the gain is not significant due to the variation of received signal strength over the frequency range [19]. It is found that the receiver gain drops by approximately 2dB at each band edge. But this drop in the gain has a negligible effect because the ten carriers at the band edge are guard carriers [1]. The receiver input matching (S11) is depicted in Figure 16 for schematic simulation and post-layout simulation. For schematic simulation, it is better than -10dB, whereas for post-layout simulation, S11 is better than -9.2dB over the entire range of 3.1GHz to 8.1GHz, as shown in Figure 16. Figure 17 depicts the IIP3 *versus* RF frequency of the receiver for both schematic and posts-layout simulations results. The IIP3 varies between -5.1dBm and -2.2dBm for schematic simulation and varies between -6.3dBm and -2.9dBm for post-layout simulation. This IIP3 is excellent for the UWB receiver to overcome cross-modulation distortions. The NF varies between 4.1dB and 7.1dB for schematic simulation as depicted in Figure 18. Figure 18 also depicts the post-layout simulation result for NF, which varies between 4.8dB and 8.92dB. The degradation in NF is due to parasitics in the post-layout extracted. This variation, which is acceptable, is due to variation of gain of the LNA. The receiver also has an excellent reverse isolation (S12) of -42dB to -59dB for schematic simulation and -32.1db to -54.2dB for post-layout simulation. The UWB receiver consumes 17.5mA at a supply voltage of 1.8V.
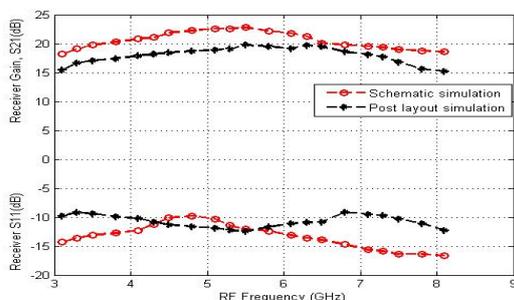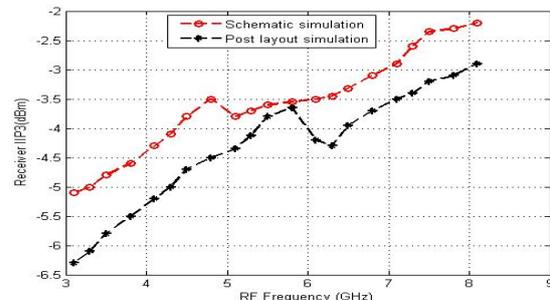

Figure 16. Receiver gain and S11.

Figure 17. Receiver IIP3 *versus* RF frequency.

Table 6 summarizes the post layout simulated performance parameters of the proposed UWB heterodyne CMOS receiver front end and comparisons with recently published similar work are also listed. It can be seen that the proposed receiver has optimized performance parameters. The comparative study further reveals that although the area of 2.55mm^2 is comparatively large due to full

203

"Fully Optimized Ultra Wideband RF Receiver Front End", R. Khatri and D. K. Mishra.

integration including all inductors, the performance parameters can still be optimized. Further work can be carried out in deep-submicron technology node like 65nm or better to further reduce the active die area. The transmitter for the UWB [20] can also be implemented in future work so that a complete transceiver can be fully integrated for UWB applications.
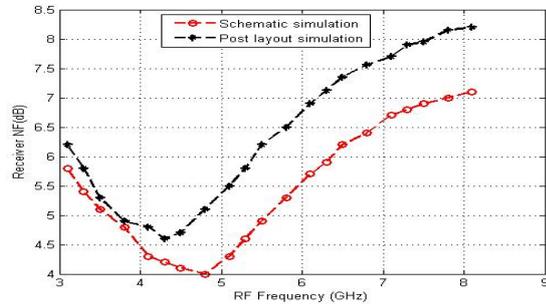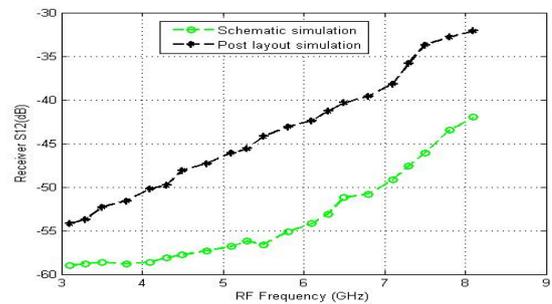

Figure 18. Receiver NF *versus* RF frequency.

Figure 19. Receiver S12 *versus* RF frequency.

Table 6. Performance comparison with recent works.

	This Work	[12]	[13]	[14]	[15]	[16]	[17]	[18]
Tech.	180nm	180nm	180nm	180nm	180nm	180nm	180nm	180nm
BW (GHz)	3.1 - 8.1	3 - 11	7.1 – 8.1	3.1 –10.6	3.1–10.6	1-6	4-10	3.1-10.6
Gain (dB)	15.2-19.8	22.8-25.8	22- 42	29	19.5-23.3	23-25	18-32	73.5
NF (dB)	4.8-8.9	4.9- 6.9	4.7	4-5.1	5.2- 9.1	2.2-2.8	3-6	8.4
IIP3 (dBm)	-6.3 to -2.9	-26	NA	-14	-10.4	-5.2 to -3.5	-6	NA
Area (mm^2)	2.55	1.04	1.43	NA	NA	3.24	2.88	3.3
Pdc (mW)	31.5	39.2	65	31.5	42	18	23	88.74

6. CONCLUSION

A 3.1GHZ – 8.1GHz CMOS UWB heterodyne receiver front end is proposed, simulated and analyzed in this paper. The proposed receiver architecture consists of a UWB LNA, a down-conversion UWB folded Gilbert mixer and an IF band pass filter. Designed and simulated in 1P6M UMC 180nm CMOS technology, the proposed receiver has optimized performance parameters as: NF of 4.8-8.9dB, gain (S21) of 15.2-19.8dB, IIP3 of -6.3dBm to -2.9dBm, consuming 17.5mA from a 1.8V supply and occupying an area of 2.55mm^2. Future research is required to design and fully integrate UWB trans-receiver for various applications.

ACKNOWLEDGEMENTS

REFERENCES

[1]   A. Batra, J. Balakrishnan, R. Aiello, J. Foerster and A. Dabak, "Design of a Multiband OFDM System for Realistic UWB Channel Environments," IEEE Trans. Microwave Theory Tech., vol. 52, no. 9, pp. 2123-2138, Sep. 2004.

[2]   C. R. Wu, H. H. Hsieh, L. S. Lai and L. H. Lu, "A 3-5GHz Frequency-tunable Receiver Frontend for Multiband Applications," IEEE Microwave and Wireless Component Letters, vol. 18, no. 9, pp. 638-640, 2008.

[3]   A. Ismile and A. Abidi, "A 3.1 to 8.2 GHz Direct Conversion Receiver for MB-OFDM UWB Communication," Proc. of the IEEE International Digest of Technical Papers. Solid-State Circuits Conference (ISSCC), pp. 208-209, San Francisco, CA, USA, 2005.

[4]   S. Stroh, "Ultra-Wideband: Multimedia Unplugged," IEEE Spectrum, vol. 40, pp. 23-27, Sep. 2003.

[5]   B. Lebung, VLSI for Wireless Communication, ISBN: 978-1-4614-0986-1, Prentice Hall, 2nd Ed, 2003.

[6]   M. Kumar, V. K. Deolia and D. Kalra, "Design and Simulation of UWB LNA Using 0.18um CMOS Technology," Proc. of the IEEE Int. Conf. on Comm., Control and Intelligent Systems (CCIS), 2016.

[7]   N. Singh, N. Yadava and R. K. Chauhan, "Design of UWB CMOS CG-LNA with Improved Noise

Figure Using Noise Cancellation Techniques," Proc. of the IEEE International Conference for Convergence in Technology (I2CT), DOI: 10.1109/I2CT45611.2019.9033747, Bombay, India, 2019.

[8] H. Seyedi, R. Dehdadt-Heydari and S. Roshni, "A Novel LNA with Noise Cancellation in 4-11.5GHz Bandwidth for UWB Receivers," Microelectronics Journal, vol.88, pp. 99-107, 2019.

[9] D. Kalra and A. Chaturvedi, "A 3.1-10.6GHz CMOS Down-conversion Mixer for UWB Systems," Proc. of the IEEE International Conference on Signal Processing & Integrated Network (SPIN), DOI: 10.1109/SPIN.2014.6777017, Noida, India, 2014.

[10] G. Bansal, A. Chaturvedi and M. Kumar, "High Linearity and High Gain Bulk Driven Down Conversion Mixer for UWB System," Proc. of the 8th IEEE Int. Conf. on Computing, Communication and Networking Technologies (ICCCNT), DOI: 10.1109/ICCCNT.2017.8204022, Delhi, India, 2017.

[11] H. Seyedi, R. Dehdasht-Heydari and S. Roshani, "UWB Down-conversion Mixer Using an IM3 Cancellation Modified Technique for Zero and Low IF Applications," Microelectronics Journal, vol. 109, pp. 1-7, 2021.

[12] Z.-Y. Huang, C.-C. Chen and C.-C. Hung, "Ultra Wideband Wireless Receiver Front-end for High Speed Indoor Applications," Journal of Engineering, vol. 2014, issue 12, pp 626-633, 2014.

[13] R. Guo and H. Zhang, "A RF Receiver Frontend for SC-UWB in 0.18um CMOS Process," Journal of Semiconductors, vol. 33, no. 12, Dec. 2012.

[14] B. Shi, M. Yan and W. Chia, "Design of a CMOS UWB Receiver Front-end with Noise-cancellation and Current-Reuse," Proc. of the IEEE International Conference on Ultra-Wideband (ICUWB2010), DOI: 10.1109/ICUWB.2010.5614618, Nanjing, China, 2010.

[15] B. Shi, M. Yan and W. Chia, "An Ultra-wideband CMOS Receiver Front-end," Proc. of the IEEE European Conf. on Wireless Technol., DOI: 10.1109/ECWT.2007.4404002, Munich, Germany, 2007.

[16] R.-F. Ye, T.-S. Horng and J.-M. Wu, "Low-noise and High Linearity Wideband CMOS Receiver Front-end Stacked with Glass Integrated Passive Decives," IEEE Transactions on Microwave Theory and Techniques, vol. 62, pp.1229-1279, May 2014.

[17] J. Manjula and S. Malarvizhi, "Active Inductor Based Tunable Multiband RF Front End Design for UWB Applications," Analog Integrated Circuits and Signal Processing, vol. 95, pp. 195-207, 2018.

[18] B. Park, K. Lee, S. Choi and S. Hong, "A 3.1-10.6 GHz RF Receiver Front-end in 0.18um CMOS for Ultra-wideband Applications," Proc. of the IEEE/MTT-S International Microwave Symposium, pp.1616-1619, Anaheim, CA, USA, 2010.

[19] N.-S. Kim and J. M. Rabaey, "A High Data-Rate Energy-Efficient Triple-channel UWB-based Cognitive Radio," IEEE Journal of Solid-State Circuits, vo. 51, no. 4, pp. 809-820, April 2016.

[20] F. R. Shahroury, "A 1.2-V Low-power Full-band Low-power UWB Transmitter with Integrated Quadrature Voltage-controlled Oscillator and RF Amplifier in 130nm CMOS Technology," Jordanian Journal of Computers and Information Technology (JJCIT), vol. 2, no. 3, pp. 165-178, December 2016.

ملخص البحث:

تقترح الورقة نهايةً أماميةً مبتكرةً ومؤمثلةً بالكامل لمستقبل ترددات راديوية فائق عرض النطاق الترددي باستخدام تقنية سي موس (IP6M 180nm UMC CMOS). والجدير بالذكر أن المعمارية الهتردودينية المستخدمة في هذا العمل لا تستخدم مازج نَبْذ الخيال المجمّع على الدارة المتكاملة. ويتألف المستقبِل المقترح من مكبّر تفاضلي منخفض الضّجيج ذي مراحل متعاقبة موصولة بنمط المنبع المشترك (CS) ومترابطة حثّياً، ومازج تحويل الى أسفل مطوي من جوع جلبرت. ويعمل المكبّر التفاضلي منخفض الضّجيج على إلغاء استخدام بالون فعّال محسِّناً الأداء المتعلّق بالضّجيج، بينما تعمل المعمارية المطوية على تقليل استهلاك الطّاقة بالنسبة للمكبّر. وقد بينت نتائج المحاكاة بعد التّصميم أنّ المكبّر لَهُ كسْب فولتية مقداره من (15.2) و(19.8) ديسيبل، ورقم ضجيج بين (4.8) و(8.9) ديسيبل، ونقطة تقاطع مدخل (IIP3) بين (6.3-) و(2.9-) dBm. كما يستهلك المكبّر ما مقداره (31.5) ميلي واط من مصدرٍ ذي فولتية قدرها (1.8) فولت. كذلك يمتلك المكبّر عزْلاً عكسيا جيداً (S12) بين (42-) و (59-) ديسيبل بسبب التركيب التعاقبي، ويحتل مساحة قدرها 2.55 ملم^2.

205

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 08, No. 02, June 2022.

RULE-BASED APPROACH FOR CONTEXT-AWARE COLLABORATIVE RECOMMENDER SYSTEM

Soulef Benhamdi[1], Abdesselam Babouri[2], Raja Chiky[3] and Jamal Nebhen[4]

ABSTRACT

Sparsity is a serious problem of collaborative filtering (CF) that has a considerable effect on recommendation quality. Contextual information is introduced in traditional recommendation systems besides users' and items' information to overcome this problem. Several research works proved that incorporating contextual information may increase sparse data. For this, data-mining techniques are among the most effective solutions that have been used in context-aware recommendation systems to handle the sparsity problem. This paper proposes the combination of a new context-user-based similarity collaborative filtering recommendation technique with data mining techniques, as a solution to this problem and develops a novel recommendation system: Rule-based Context-aware Recommender System (R_CARS). R_CARS is experimented introducing four rule-based algorithms: JRip, PART, J48 and RandomForest, on four different datasets: DePaulMovie, InCarMusic, Restaurant and LDOS_CoMoDa and compared with the state-of-the-art models. The results of the experiment show that weighting the rating-based similarity with context and combining it with a rule-based technique can overcome the sparsity problem and significantly improve the accuracy of recommendation compared to the state-of-the-art models.

KEYWORDS

1. INTRODUCTION

Recommendation systems (RSs) are considered as the most effective solution to the problem of information overload in online systems [1]. RSs use information on users and items for providing recommendations to the target user. Three main approaches are used by this kind of system [2]-[3]:

- Content-based filtering generates recommendations to a user with regard to his/her preferences in the past.
- Collaborative filtering models are built from collected user-item interactions. By analyzing behavioural patterns across the whole user base, the recommendations are based on extracting interactions similarities among users, items or both.
- Hybrid recommender systems combine various methods, exploiting the benefits of each one.

Recent research indicated that users' preferences change according to the context [4]. For example, a user may choose a different kind of movie if he or she is going to watch the movie at *home* rather than at the *cinema*. Therefore, contextual factors are introduced to alleviate some traditional RS issues, especially sparsity [5] and consequently enhance the recommendation accuracy. In fact, incorporating contextual information may not alleviate the sparsity problem, but may reduce its effect by reducing sparse data, because " it is not guaranteed, under the same contextual situation, that users rate all the items" [6]. For this, several approaches are suggested to face sparsity problems in context-aware recommendation systems (CARSs): selection of influential contextual attributes [6], similarity-based collaborative filtering [7], data-mining techniques: sequencing patterns, classification, association, regression, Computational Intelligence (CI) techniques, …etc. [8]-[13].

There are three main techniques used to develop a context-aware recommendation system according to the step where the contextual information is integrated [14]: *pre-filtering, post-filtering and contextual modeling*. In the *pre-filtering* technique [15] (Figure 1 (a)), R represents users' ratings, C represents contextual information and contextual information is used as input data with ratings to generate recommendations; i.e., irrelevant ratings are filtered out using contexts. Afterward, the recommendation model is built using classic recommendation approaches. *Post-filtering* (Figure 1 (b)) [16], one of the

1. S. Benhamdi is with Dept. of Computer Sci., Uni. 8 Mai 1945, Guelma, Algeria. Email: benhamdi.soulef@univ-guelma.dz
2. A. Babouri is with LGEG Laboratory, Uni. 8 Mai 1945, Guelma, Algeria. Email: babouri.abdesselam@univ-guelma.dz
3. R. Chiky is with ISEP, Paris, France. Email: raja.chiky@isep.fr
4. J. Nebhen is with College of Computer Sci. and Eng., Prince Sattam bin Abdulaziz Uni., KSA. Email: j.nebhen@psau.edu.sa

classic recommendation models, is used to predict ratings. Afterward, contextual data is used to adjust the predicted ratings. *Contextual modeling* technique (Figure 1 (c)) [17] incorporates contextual information in the ranking function to calculate predictions.

The contributions of this paper are to;
- Design a pre-filtering context-aware recommender system framework: Rule-based Context-aware Recommender System (R_CARS), that is based on context-user-based similarity collaborative filtering and a rule-based technique. The rating-based similarity is weighted using context-based similarity.
- Alleviate data sparsity using a rule-based approach with a context-based similarity technique.
- For comparative raisons, an experiment is conducted with the most popular classifiers: ZeroR, OneR, REPTree, RandomTree, JRip, PART, J48 and RandomForest. The four last classifiers are selected based on their performance in term of accuracy to be presented in this paper. The proposed model is compared with the state-of-the-art models.

The first section gives an overview of recommendation systems and research aims. Section two presents related work, whereas section three details the proposed approach. Section four presents the experiment results. Section five concludes this paper.

2. RELATED WORK

Context-awareness in recommendation systems is a novel research domain that attracted researchers in the last decade [18]-[19]. In real life, the users' choices may vary from one situation to another; for example, time, companion and location are important factors that users may take into account when selecting a movie [20]. The user's activity (e.g. driving) or emotional attributes (e.g. mood) may affect his/her choice when listening to music [21]. Location, time, type of food served, price of the meal…, might be important factors that should be involved in recommending services (hotels, restaurants, …etc.) to tourists [22].

This paper focuses on research works that exploit context-similarity and/or rule-based approaches to predict ratings. To face the sparsity and cold start problems, a recommendation algorithm (called TCAR) that uses association rule mining and an improved overlapping community detection method is proposed in [23]. Time-weighted similarity between users is computed to detect overlapping communities. Association rule mining is employed to model users' drifts over time. However, this method takes into consideration only a single context attribute. [24] suggests a rule-based recommendation method to predict ratings. To reduce sparse data, this method aims to reduce the number of context features; i.e., construct low-dimensional latent contexts. [25] proposes a context similarity measure within a multi-criteria collaborative recommendation approach for service recommendation. To prevent sparsity, ratings are used within similar contexts. A ranking-based recommendation algorithm (called SLIM for Sparse Linear Method) is combined with context similarity in [26] to handle the sparsity problem. A context-aware location recommendation for groups based on the random walk (CLGRW) algorithm is proposed in [27]. Time and weather conditions are also introduced. This approach exploits content similarity and location popularity to reduce sparse data and improve the performance of recommendation. A context-aware collaborative filtering and a knowledge-based approach are combined in [28] to construct a context-aware recommender system for an m-tourism application. This recommendation model provides to a tourist, using a personal mobile device, recommendations on attractions, restaurants, hotels and transportation with regard to a current situation and its profile. [29] proposes a context-aware recommendation approach that uses a combination of novel graph-based similarity with nearest neighbor methods to face the negative effect of sparsity. [4] proposes a novel recommender system, called RSTRC (Recommender System based on Temporal Reliable and Confidence measures), incorporating time factor into reliable and confidence measures. A probabilistic approach is used to evaluate the effectiveness of users' rating profiles and an enrichment mechanism is applied to deal with the sparsity problem by making a denser user-item rating matrix. A temporal similarity measure is used to compute the similarity values of users. [30] proposes two models: CUBCF (Context-similarity User-based Collaborative Filtering recommender model) and CIBCF (Context-similarity Item-based Collaborative Filtering recommender model), using a context-similarity collaborative filtering approach that is exploited to alleviate the sparsity problem of CF models and increase their efficiency. CUBCF is based on user-based collaborative filtering. Split techniques are used to compute the rating-based similarity and chi square kernel is used to compute context-based

207

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 08, No. 02, June 2022.

similarity. Then, the similarity among users is computed using rating-based and the context-based similarity. In the final step, CUBCF recommends to the active user items with the highest similarity value. CIBCF is based on item-based collaborative filtering. Split techniques are used to compute rating-based similarity and chi square kernel is used to compute context-based similarity. Then, the similarity among items is computed using rating-based and context-based similarity. In the final step, CIBCF recommends items with the highest similarity value. [31] proposes a CARS scheme: CACF (Context-aware Collaborative Filtering) that is based on pre-filtering and post-filtering paradigms with context-weighting approach using Real-coded Genetic Algorithm (RCGA). Further, an Effective Missing Value Prediction (EMPV) algorithm [32] is adopted to handle the sparsity problem. In [33], a non-dominated user neighborhood concept is introduced into DCM (Differential Context Modeling) approaches [34] to develop a CARS framework: ND-DCM (Non-dominated Differential Context Modeling). The non-dominated user neighborhood calculates user-user similarities from multiple dimensions; i.e., the selected similar users have higher similarities with the target user than others from multiple perspectives (e.g. ratings, demographic information, social relationships, …etc.). The DCM approach proposes two context aware recommendation models based on UBCF (User-based Collaborative Filtering): Differential Context Relaxation (DCR) and Differential Context Weighting (DCW). To alleviate the sparsity issue, DCR exploits the optimal-context matching on the rating profiles, while DCW utilizes a weighted context similarity. ND-DCM proposes four approaches: ND-DCR (Non-dominated Differential Context Relaxation), NDs-DCR, ND-DCW (Non-dominated Differential Context Weighting) and NDs-DCW. ND-DCR and ND-DCW measure, respectively, user-user similarities and weighted user similarities based on two matrices: user-item rating matrix (UI) and user-condition rating matrix (UC). NDs-DCR and NDs-DCW measure user similarities based on single rating matrix (UC) with conditions associated with all the selected context dimensions.

Almost all the above mentioned methods are specialized in a specific domain (movies, music, tourism, …etc.), using specific contextual attributes, except CUBCF and the model proposed in [29] which could be applied in any domain. Thus, all the state-of-the-art methods, including these two last models, use splitting techniques to define the target user's neighbors. In this work, similarity is computed in a more simple way; i.e., the similarity is calculated between the target user and all the other users. Then, the users are ordered according to the similarity values and the n first users are defined as the most similar users. Since the authors adopted a user-based approach, R_CARS is compared to the models: CUBCF, ND-DCR, NDs-DCR, ND-DCW and NDs-DCW. The main difference between R_CARS and the other models is that R_CARS adopts a rule-based approach to predict rating, while these models use only context similarity. The objective of this comparison is to illustrate the impact of using rule-based recommendation with a similarity-based collaborative approach on recommendation efficiency.

3. RULE-BASED CONTEXT-AWARE RECOMMENDER MODEL (R_CARS)

In this section, the proposed recommendation model: rule-based context-aware recommender system (R_CARS), is detailed. R_CARS is built based on context-aware collaborative filtering recommendation and rule-based recommendation approach. First, the dataset is preprocessed. Second, R_CARS calculates similarities between the target user and the other users. Finally, the rule-based algorithm is applied on the similar users' dataset in order to generate recommendations (Figure 2).
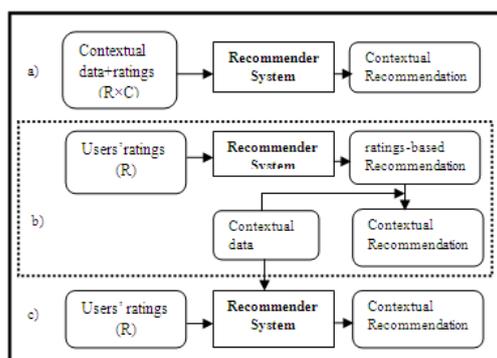

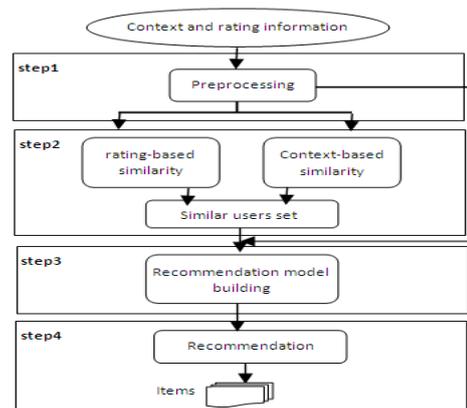Figure 1. Context-aware recommender models: (a) Pre-filtering, (b) Post-filtering, (c) Contextual modeling.

Figure 2. The R_CARS model.

3.1 System Model

The system model used in this work is:
- Let $U=\{u_0,u_1,\ldots u_n\}$, a set of users.
- Let $I=\{i_1,i_2,\ldots i_w\}$, a set of items.
- $R_a=\{r_{a1},r_{a2},\ldots r_{am}\}$ a set of m ratings given by user u_a, with a set of k contextual attributes $C_a=\{c_0,c_1,\ldots c_x\}$. $R_a \subset R$, R is the set of all ratings.
- $P=\{p_{a1}, p_{a2},\ldots p_{aq}\}$ is a set of m predicted ratings to the target user a.
- $L=\{l_1, l_2,\ldots l_z\}$ is a set of induction rules.

3.2 Datasets

In this paper, four real-world datasets are used: DePaulMovie, TheCarMusic, LDOS_CoMoDa and Restaurant.

- DePaulMovie [35]: This dataset contains 5043 ratings provided by 97 users on 319 films, based on three contextual attributes: time, companion and location (the rating scale 1-5).

- InCarMusic [36] contains 4012 ratings provided by 43 users on 139 songs, based on eight contextual attributes: driving style, mood, landscape, …etc. (the rating scale 0-5).

- Restaurant [37] is a public dataset that contains 50 subjects who gave ratings (the rating scale 1-5) to 40 popular restaurants in Tijuana. There are 6 contexts in total which are the combinations of the original two contextual attributes: time and location. In the data, the contexts are represented by S1-S6: S1 (Weekday+school), S2 (Weekend+home), S3 (Weekday+work), S4 (Weekend+school), S5 (Weekend+home), S6 (Weekend+work).

-LDOS_CoMoDa [38] is a public dataset that contains 121 users who rated 1232 movies considering 12 context dimensions (mood, weather, time, location, …etc.). The rating scale 1-5.

3.3 Data Preprocessing

Preprocessing data is the first step in machine learning. The data could contain errors or anomalies that may affect the analysis process, such as: null values, irrelevant features or attributes, inappropriate feature type …etc. Therefore, the data should be preprocessed before applying machine-learning algorithms: classification, regression, clustering, …etc.

Unnecessary attributes are removed, which should not be involved in the recommendation process, such as user Id and item Id. Some other nominal feature types are converted into a string (applying the unsupervised filter: *NominalToString*) before similarity computing. After similarity computing, we reconvert them into nominal (applying the unsupervised filter: *StringToNominal*). Numeric attributes are also converted into nominal, in order to apply the predictive algorithms accurately (using Weka API). For example, time, location, landscape, driving style, rating…etc. In order to improve the prediction results, unsupervised discretization filter (*Discretize*) is used to descritize "rating" into three intervals (categories or classes). Each interval or class presents a rating category. Three categories are defined: the first category C1: [0-2,33] represents non-interesting items; i.e., items having low ratings, while the second category C2:[2,33-3,66] represents items rated with medium values and the third category C3:[3,66-5] represents interesting items (items rated with high values).

3.4 Similarity Computing

Almost all collaborative recommender systems use split techniques in order to define the target user's similar users. These techniques split users into clusters, calculate the distance between the target user and each cluster and then assign the target user to the appropriate cluster. The proposed technique is different and simpler; similarity is computed between the target user and all the other users. Then, the users are ordered according to the similarity values and the n first users are defined as the most similar users. To give flexibility to the experiments, n is selected as 10:5:25 empirically. To calculate the similarity between two users, the authors propose the definition of similarity based on their ratings and similarity based on contextual attributes when these ratings are given. Each user's profile is defined by a matrix with rows as items and columns as ratings and contextual attributes. Each row is composed of ratings and contextual values given on the correspondent item. The users' u_a and u_b profiles are defined

209

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 08, No. 02, June 2022.

as follows:

$$u_a = \begin{pmatrix} r_{a,1} & c1_{a,1} & c2_{a,1} & ck_{a,1} \\ r_{a,2} & c1_{a,2} & c2_{a,2} & ck_{a,2} \\ r_{a,m} & c1_{a,m} & c2_{a,m} & ck_{a,m} \end{pmatrix} \tag{1}$$

$$u_b = \begin{pmatrix} r_{b,1} & c1_{b,1} & c2_{b,1} & ck_{b,1} \\ r_{b,2} & c1_{b,2} & c2_{b,2} & ck_{b,2} \\ r_{b,m} & c1_{b,m} & c2_{b,m} & ck_{b,m} \end{pmatrix} \tag{2}$$

where $r_{a,m}$ is the rating given by a user u_a on item i_m. The similarity between u_a and u_b is calculated taking into account the items that are co-rated by u_a and u_b. For example, if u_a and u_b rated the items i_1 and i_2, similarity is calculated based on the ratings and contextual values given by the two users on i_1 and i_2. Here, all context attributes are converted from nominal into string for implementation raisons. '*Rating*' is still unchanged (numeric). Formula (3) is proposed to calculate similarity.

$$sim(u_a, u_b) = \sum_{j=1}^{k}(r_{aj}-r_{bj})^2 \times (nb_{mat}/nb) \tag{3}$$

where *sim(u_a, u_b)* calculates the similarity between u_a and u_b. To define ratings-based similarity between u_a and u_b, the sum of the squared differences of their ratings is calculated. r_{aj} is the rating value given by user u_a on item j. r_{bj} is the rating value given by user u_b on item jϵ I. s is the number of items co-rated by users u_a and u_b. The second term calculates similarity based on contextual attributes. The context-based similarity between u_a and u_b is the ratio between the number of match contextual attributes values and the number of available contextual attribute values. nb_{mat} is the number of common contextual attribute values related to items that have been co-rated by users u_a and u_b, whereas nb is the number of all contextual attribute values that user a (or user b) should give. Figure 3 shows how similarity between an active user u_0 and the users: u_1, u_2 and u_3 is computed.

3.5 Rating Prediction

The proposed recommender system is based on one of the data mining techniques: classification. Data mining, also called Knowledge Discovery in Databases (KDD) [39], is the process of discovering interesting information, previously unknown, in a large amount of data involving techniques from machine learning, statistics and database systems [40]. Classification, also called supervised learning, is a data-mining task that assigns items in a dataset to target categories or classes [41, 43]. The first step of classification is the model construction using the training set, which is divided into predefined classes. The model could be built as classification rules, decision trees or mathematical formulae. Model exploitation is the second step, in which new objects from the test dataset, which is independent of the training dataset, will be classified and then the results are compared to the previous step results in order to evaluate the model accuracy [43]. A rule-based classification refers to any classification scheme that makes use of "IF-THEN" rules to predict the class for new items. ZeroR, OneR, PART, JRip, DTNB, ConjunctiveRules, M5Rule, Prism [39]… are examples of rule-based classifiers. Decision trees represents a graph-based classification method that uses tree-like structure to make decisions. A tree is composed of: root node, internal nodes that represent conditions (applied on attributes) and leaf nodes, where leaf nodes' contents represent outcomes or classes. The path between the root node and one of the leaf nodes presents a rule. The rules in decision trees have the form: If (condition1) and (condition 2) and (condition3)…Then outcome [44]. J48, RandomForst, REPTree, RandomTree, NBTree, ADTree, DecisionStump, BFTree… are examples of decision trees.

After identifying the similar users set, data preprocessing is necessary (converting string and numeric attributes into nominal). Then, R_CARS utilizes a classification algorithm to predict ratings (e.g. J48). Next, we explain how predictions are computed using J48.

3.5.1 J48 Classifier

J48 is an open-source Java implementation of the statistical classifier: C4.5 [45]. C4.5 generates a trimmed decision tree from a set of training data and in the same way, the ID3 algorithm does, by using the concept of information entropy. Entropy is a measure of the amount of uncertainty in the dataset *D*. Mathematical representation of entropy is:

$$H(D) = \sum_{cl_i \in CL} -p(cl_i)log_2\, p(cl_i) \tag{4}$$

- D is the current dataset for which entropy is being calculated.
- CL is a set of classes in D.
- $p(cl_i)$ is the proportion of the number of elements in class cl_i to the number of elements in set D.

The entropy allows the computing of the information gain of each attribute. Information gain (a) tells us how much uncertainty in D was reduced after splitting set D on attribute a. Mathematical representation of information gain is:

$$IG(a, D) = H(D) - \sum_{i=1}^{v} p(d_i)H(d_i) \tag{5}$$

- d_i is the subset created from splitting set D by attribute a ($D=\{d_1,\ldots,d_v\}$).
- $p(d_i)$ is the proportion of the number of elements in subset d_i to the number of elements in set D.
- $H(d_i)$ is the entropy of subset d_i.

The J48 algorithm's steps can be defined as follow:

- First, select the attribute with the greatest information gain for the root node and create a new child node for each possible value.
- Then, split the training set into child nodes (subsets).
- Repeat recursively for each child node until getting pure subsets.

3.5.2 J48 for Prediction Computing

In this sub-section, we explain how predictions are calculated using an example from Restaurant Dataset (Figure 4). This example presents the data provided by three similar users. The *rating* attribute is defined as the class attribute. After calculating the information gain of each attribute, the decision tree is built. Then, a set of rules are generated L. To make recommendation for an active user u_0, it is necessary to go down the decision tree and apply the appropriate rule. We suppose that the active user u_0 selects '*applebees*' restaurant under Context S2 (Weekend+home); so, the classifier will predict the rating category: C1, applying the rule l_1.

3.5.3 Alleviating Sparsity

Each data-mining technique has its internal strategy to handle sparsity: Distribution-based Imputation (DBI), Unique Value Imputation (UVI), Predictive Value Imputation (PVI), …etc. [46]. For example, decision trees (C4.5 algorithm) use Distributed-based Imputation (DBI) method to predict missing values [10]. DBI splits the instance to be classified into sub-instances each with different values for missing feature and weight corresponding to the estimated probability of the particular missing value. Each sub-instance is routed down the appropriate tree branch according to its assigned value. Upon reaching a leaf node, the class-membership probability of the sub-instance is measured as the frequency of the class in the training instances associated with this leaf. The overall estimated probability of class membership is calculated as the weighted average of the class membership probabilities over all sub-instances. If there is more than one missing value, the process recurses with the weights combining multiplicatively [46]. For example, suppose attribute a has two possible values: a_1 and a_2; if a node contains 6 branches with $a(?)=a1$ and 4 with $a(?)=a2$, then the probability that $a(?)= a1$ is 0.6 and the

Figure 3. Similarity computing.
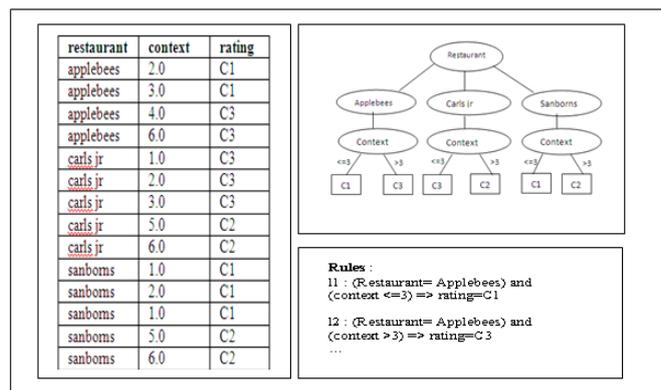
Figure 4. An example of a decision tree (J48).

211

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 08, No. 02, June 2022.

probability that *a(?)=a2* is 0.4. A fractional 0.6 of (*?*) is now distributed down the branch for a=a1 and a fractional 0.3 of (*?*) down the other tree branch.

3.6 R_CARS Algorithm

The following algorithm summarizes the recommendation model steps:

R_CARS Algorithm

Input: R, I, a given target user u_a.
Output: top k items with highest prediction score to u_a.
Begin
Step 1: Data preprocessing.
Step 2: - Calculate the similarity between u_a and each user of the initial dataset by Eq. (3).
     - Construct the top n (n=10:5:25) similar users' dataset.
Step 3: - Data preprocessing.
    - Build the recommendation model using the similar users' dataset.
 Step 4: **- G**enerate predictions.
    - Return the top k items.
 End.

4. EXPERIMENT AND FRAMEWORK

To evaluate its effectiveness, R_CAR is implemented and experimented using Weka JAVA API, applying 5-fold cross-validation, on the context-aware datasets: DePaulMovie, InCarMusic, Restaurant and LDOS_CoMoDa, where InCarMusic contains more sparse data than DePaulMovie and LDOS_CoMoDa contains more sparse data than Restaurant. Two experiments are conducted:

- Experiment 1: R_CAR is experimented on DePaulMovie and InCarMusic datasets, introducing four classifiers (JRip, PART, J48 and RandomForest), in order to select the best one compared with CUBCF.

- Experiment 2: R_CAR is experimented on Restaurant and LDOS_CoMoDa datasets, introducing the four classifiers, in order to select the best one compared with ND-DCR, NDs-DCR, ND-DCWand NDs-DCW models.

4.1 Evaluation Metrics

4.1.1 Deviation-based Evaluation Metrics

These metrics, also known as error-based metrics, measure the deviation between the predicted rating value and the real rating value to evaluate the performance of the recommendation algorithm. The deviation-based metrics are: MSE (Mean Squared Error), RMSE (Root Mean Squared Error), MAE (Mean Absolute Error), MAPE (Mean Absolute Percentage Error) and SMAPE (Symmetric Mean Absolute Percentage Error) [47]. Here, MAE and RMSE are used to evaluate the performnce of R_CARS.

$$\text{MAE} = \sum_1^S |raj - paj|/S \tag{6}$$

$$\text{MSE} = \sum_1^S (raj - paj)^2/S \tag{7}$$

$$\text{RMSE} = \sqrt{\left(\sum_1^S (raj - paj)^2/S\right)} \tag{8}$$

$$\text{MAPE} = 1/S \sum_1^S |raj - paj/raj| \tag{9}$$

$$\text{SMAPE} = 1/S \sum_1^S 2|raj - paj|/(|raj| + |paj|) \tag{10}$$

r_{aj} is the user's a explicit rating to item j. p_{aj} is the user a's predicted rating to item j. S is the number of predicted ratings. The low values of these metrics indicate that the recommendation model is good [48]-[49].

4.1.2 List-based Evaluation Metrics

Lists-based metrics, also known as TOP k recommendation evaluation metrics, verify whether the recommender system is able to recommend interesting items by calculating the rate of interesting items

among those recommended; that is to evaluate the efficiency of the recommendation algorithm using: Precision, Recall and F-measure [48]-[49].

$$\text{Precision} = TP/TP + FP \qquad (11)$$

$$\text{Recall} = TP/TP + FP \qquad (12)$$

$$F_{measure} = 2 * \text{Precision} * \text{Recall}/\text{Precision} + \text{Recall} \qquad (13)$$

- TP (True Positives): interesting items recommended by the recommendation model.
- FP (False Positives): uninteresting items recommended by the recommendation model.
- TN (True Negatives): uninteresting items not recommended by the recommendation model.
- FN (False Negatives): available interesting items not recommended by the recommendation model.

4.1.3 Classification Evaluation Metric

In machine learning, the accuracy metric is used to measure the performance of classification algorithms, By comparing the results of the classifier building using the training dataset with the testing classifier results, Accuracy calculates the percentage of items correctly classified [9]. Accuracy is used to evaluate the performance of the four classifiers introduced in this work.

4.2 Experiment 1

The recommendations are presented to an active user u0. The performance of R_CARS is measured using the metrics: MAE, RMSE, Precision, Recall, F-measure and accuracy. These metrics are computed according to the number of similar users (n=10:5:25). The average for each classifier on DePaulMovie and InCarMusic datasets is presented in Table 1 and Table 2, respectively (only the average is presented in this work). In [30], the proposed model: CUBCF, is experimented on the two datasets: DePaulMovie and InCarMusic, applying 5-fold cross-validation and compared with UBCF, IBCF and CIBCF (The results of this experiment are shown in Tables 1 and 2 [30]). A minimal similarity threshold is set to select neighbours.

4.3 Experiment 2

In this experiment, we used the same parameters' values as in the previous experiment (n=10:5:25). Table 3 and Table 4 show the evaluation metrics' values (average) for each classifier on Reastaurant and LDOS_CoMoDa, respectively. In [33], the proposed models: ND-DCR, NDs-DCR, ND-DCWand NDs-DCW, are experimented on the two datasets: Restaurant and LDOS_CoMoDa, applying 5-fold cross-validation (The results of this experiment are shown in Tables 3 and 4 [33]) and compared to some baseline recommendation algorithms: DCW, DCR, CAMF (Context-aware Matrix Factorization), DCW-GA (DCW Genetic Algorithm), …etc. A minimal similarity threshold is set to select neighbours. Only MAE and RMSE are used to compare R_CARS with these models.
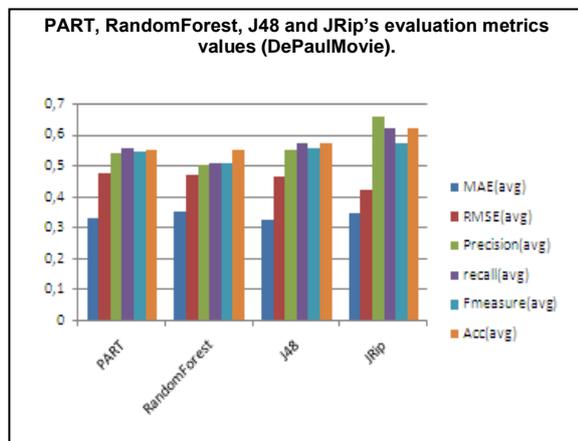

Figure 5. PART, RandomForest, J48 and JRip evaluation metrics' values (DePaulMovie).
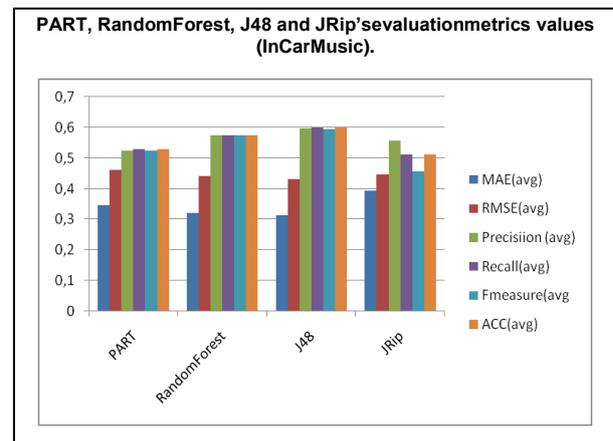

Figure 6. PART, RandomForest, J48 and JRip evaluation metrics' values (InCarMusic).

213

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 08, No. 02, June 2022.

Table 1. Evaluation metrics' values of the four classifiers (DePaulMovie).

Models	MAE	RMSE	Precision	Recall	F-measure	ACC(%)
R_CAR(PART)	0,327775	0,4713	0,53625	0,55225	0,5425	0,55185
R_CAR(RandomF)	0,348275	0,467375	0,501	0,50775	0,504	0,5075
R_CAR(J48)	**0,3251**	**0,46265**	**0,55**	**0,57025**	**0,55725**	**0,57035**
R_CAR(Jrip)	0,3429	0,4203	0,6595	0,6195	0,571	0,61945
CUBCF	1,39157	1,94572	<0,50	<0,25	<0,25	/

Table 2. Evaluation metrics' values of the four classifiers (InCarMusic).

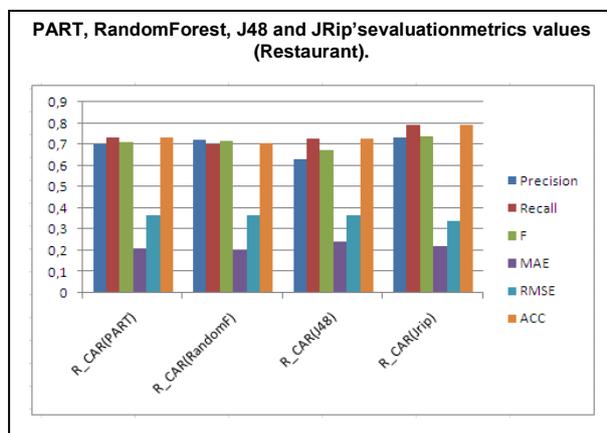| Models	MAE	RMSE	Precision	Recall	F-measure	ACC(%)
R_CAR(PART)	0,347075	0,460625	0,52375	0,5295	0,52375	0,5291
R_CAR(RandomF)	0,3203	0,44065	0,5735	0,5735	0,5735	0,5747
R_CAR(J48)	**0,31255**	**0,432175**	**0,597**	**0,60075**	**0,59575**	**0,6005**
R_CAR(Jrip)	0,392725	0,44685	0,558	0,513	0,45675	0,51295
CUBCF	1,058	1,28781	< 0,2	< 0,05	< 0,1	**/**

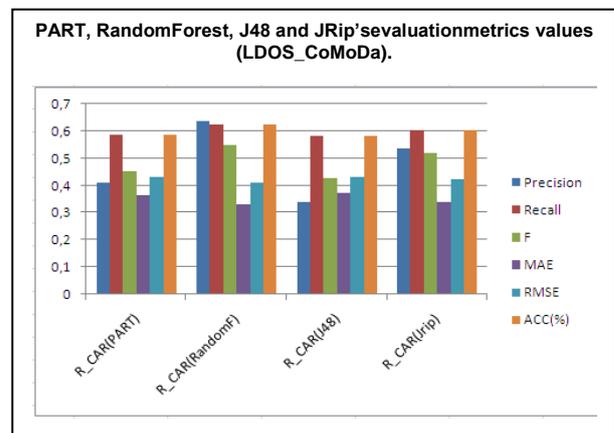Figure 7. PART, RandomForest, J48 and JRip evaluation metrics' values (Restaurant).

Figure 8. PART, RandomForest, J48 and JRip evaluation metrics' values (LDOS_CoMoDa).

Table 3. Evaluation metrics' values of the four classifiers (Restaurant).

Models	MAE	RMSE	Precision	Recall	F-measure	ACC(%)
R_CAR(PART)	0,210325	0,365425	0,698	0,735	0,71225	0,73475
R_CAR(RandomF)	0,206425	0,36645	0,72475	0,7025	0,715	0,706925
R_CAR(J48)	0,2451	0,364925	0,62975	0,7295	0,6715	0,729675
R_CAR(Jrip)	**0,218625**	**0,340675**	**0,73225**	**0,792**	**0,7365**	**0,792**
ND-DCR	0,784	1,090	/	/	/	/
NDs-DCR	0,787	1,058	/	/	/	/
ND-DCW	0,739	1,002	/	/	/	/
NDs-DCW	0,735	0,997	/	/	/	/

Table 4. Evaluation metrics' values of the four classifiers (LDOS_CoMoDa).

Models	MAE	RMSE	Precision	Recall	F-measure	ACC(%)
R_CAR(PART)	0,36485	0,431725	0,40975	0,58725	0,45375	0,5871
R_CAR (RandomF)	**0,332425**	**0,410875**	**0,63625**	**0,62325**	**0,5495**	**0,6232**
R_CAR(J48)	0,3743	0,430125	0,33875	0,5815	0,428	0,58135
R_CAR(Jrip)	0,34145	0,421425	0,53775	0,60475	0,521	0,60465
ND-DCR	0 ,723	0,939	/	/	/	/
NDs-DCR	0,726	0,938	/	/	/	/
ND-DCW	0,731	0,927	/	/	/	/
NDs-DCW	0, 714	0,927	/	/	/	/

4.4 Result Discussion

In this sub-section, the results obtained in experiment 1 and experiment 2 are discussed. When using DePaulMovie (Table 1), we can note that J48 is the best performing algorithm, having lower error indicators' values and higher values for Precision, Recall, F-measure and accuracy. Most of the results between J48 and JRip are not significant (taking into account all the evaluation metrics). We can also note that R_CARS, introducing the four classifiers, is more efficient than CUBCF (MAE: the difference is over 1,0, RMSE: the difference is over 1,5, Precision: the difference is approximately 0,1, Recall and F: the difference is over 0,3).

Using sparser dataset: InCarMusic (Table 2), we note that J48 outperforms the other algorithms. J48 has lower error indicators' values and higher values of Precision, Recall, F-measure and accuracy, which means that J48 is the best even on datasets with a significant number of sparse data. The results between J48 and RandomForest are not significant. R_CARS, using the four classifiers, can beat CUBCF model (MAE: the difference is over 0,7, RMSE: the difference is over 0,8, Precision: the difference is over 0,3, Recall: the difference is approximately 0,5, F: the difference is over 0,4). CUBCF performance decreases significantly on InCarMusic dataset. This is not surprising since CUBCF depends only on context-based similarity to alleviate sparsity.

When using Restaurant dataset (Table 3), we can see that JRip is the best performing algorithm. The results between JRip and RandomForest are not significant. R_CARS, introducing the four classifiers, can beat the models: ND-DCR, NDs-DCR and ND-DCW (MAE: the difference is over 0,5, RMSE: the difference is approximately 0,7).

Using sparser dataset: LDOS-CoMoDa (Table 4), the experimental results illustrate that RandomForest outperforms the other algorithms, which means that RandomForest is the most efficient on datasets with high level of sparsity. The results between JRip and RandomForest are not significant. Although R_CARS's (introducing the four classifiers) performance decreases, it is still more efficient than those of ND-DCR, NDs-DCR, ND-DCW and. ND-DCW (MAE: the difference is over 0,4, RMSE: the difference is over 0,5). Recall that these models utilize the non-dominated user neighbourhood to alleviate sparsity, while R_CARS combines context-based similarity with data-mining techniques to face this problem.

As a summary, the results demonstrate that RandomForest works better than the other classifiers on datasets with high number of sparse data, while JRip works better on less sparse datasets. In comparison with the state-of-the-art models, R_CARS, taking into consideration the four classifiers, is still the most efficient on the four datasets.

5. CONCLUSIONS

CARSs exploit the knowledge about context under which the ratings are given to generate recommendation. However, in this case, the sparsity issue could be more serious, since there may be limited ratings within the same contextual situation, especially when there are many context dimensions. This paper proposes a novel context-aware recommendation model that we called R_CARS to overcome sparsity issues by combining the proposed user-based similarity technique with a rule-based approach. R_CARS is experimented on four real-world datasets introducing four rule-based algorithms: PART, JRip, RandomForest and J48 and compared to the state-of-the-art models.

The experimental results demonstrate that RandomForest works better than the other classifiers on datasets with a high number of sparse data, while JRip works better on less sparse datasets. The results prove also that R_CARS, introducing the four classifiers, can alleviate the sparsity problem and provide more accurate recommendations compared to the other models. In future work, we will compare the proposed similarity technique with state-of-the-art similarity techniques by using large datasets. We will also work on implementing more effective solutions to overcome the sparsity problem by taking advantage of recent approaches, such as deep-learning techniques [50]-[51].

REFERENCES

[1]   S. Benhamdi, A. Babouri and R. Chiky, "Personnalized Recommender System for E-learning Environment," Education and Information Technologies, vol. 22, no. 4, pp. 1455–1477, 2017.

[2]   J. Bobadilla, F. Ortega, A. Hernando and A. Gutiérrez, "Recommender Systems Survey," Knowledge-

215

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 08, No. 02, June 2022.

based Systems, vol. 46, pp. 109–132, 2013.

[3]   A. S. Ghabayen and B. H. Ahmed, "Enhancing Collaborative Filtering Recommendation Using Review Text Clustering," Jordanian Journal of Computers and Information Technology (JJCIT) ,vol. 7, no. 2, pp. 152 - 165, DOI: 10.5455/jjcit.71-1609969782, June 2021.

[4]   S. Ahmadian, N. Joorabloo, M. Jalili and M. Ahmadian, "Alleviating Data Sparsity Problem in Time-aware Recommender Systems Using a Reliable Rating Profile Enrichment Approach," Expert Systems with Applications, vol. 187, p. 115849, 2022.

[5]   G. S. Eshwari and S. P. S. Ibrahim, "Rule-based Effective Collaborative Recommendation Using Unfavourable Preference," IEEE Access, vol. 8, pp. 128116 -128123, 2020.

[6]   S. M. Abbes, K. A. Alam and S. Shamshirband, "A SoftrRough Set Based Approach for Handling Contextual Sparsity in Context-aware Video Recommender Systems," Mathematics, vol. 7, no. 8, p. 740, 2019.

[7]   Z. Huang, X. Lu and H. Duan, "Context-aware Recommendation Using Rough Set Model and Collaborative Filtering," Artificial Intelligence Review, vol. 35, pp. 85- 99, 2011.

[8]   M. Jamali and M. Ester, "Mining Social Networks for Recommendation," Tutorial of ICDM, vol. 11, [Online], Available: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.367.4838&rep=rep1&type=pdf, 2011.

[9]   T. Sanli, C. Sicakyuz and O. H. Yuregir, "Comparison of the Accuracy of Classification Algorithms on Three Datasets in Data Mining: Example of 20 Classes," International Journal of Engineering, Science and Technology, vol. 12, no. 3, pp. 81-89, 2020.

[10]   S. F. Huang and C. H. Cheng, "A Safe-region Imputation Method for Handling Medical Data with Missing Values," Symmetry, vol. 12, no. 11, p. 1792, 2020.

[11]   E. Kolce and N. Frasheri, "A Literature Review of Data Mining Techniques Used in Healthcare Databases," Proc. of the ICT Innovations, Web Proceedings, pp. 577-582, 2012.

[12]   V. Patil and V. B. Nikam, "Study of Data Mining Algorithm in Cloud Computing Using MapReduce Framework," Journal of Engineering, Computers & Applied Sciences (JEC&AS), vol. 2, no. 7, pp. 65-70, 2013.

[13]   E. Osmanbegović and M. Suljić, "Data Mining Approach for Predicting Student Performance," Economic Review – Journal of Economics and Business, vol. X, no. 1 , pp. 3-12, May 2012.

[14]   I. F. Tobias, P. G. Campos, I. Contador and F. Dies, "A Contextual Modeling Approach for Model-based Recommender System," Proc. of the Conference of Spanish Association for Artificial Intelligence, vol. 8109, pp. 42-51, DOI: 10.1007/978-3-642-406432-0-5, 2013.

[15]   V. Codina, F. Ricci and Luigi Ceccaroni, "Distributional Semantic Pre-filtering in Context-aware Recommender Systems," User Modeling and User Adapted Interaction, vol. 26, pp.1–32, 2015.

[16]   X. Ramirez-Garcia and M. Garca-Valdez, "Post-filtering for a Restaurant Context-aware Recommender System," Chapter in Book: Recent Advances on Hybrid Approaches for Designing Intelligent Systems, Part of the Studies in Computational Intelligence Book Series, vol. 547, pp. 695–707, 2014.

[17]   Y. Zheng, "Deviation-based and Similarity-based Contextual SLIM Recommendation Algorithms," Proc. of the 8th ACM Conference on Recommender Systems (RecSys '14), pp. 437-440, DOI:10.1109/ACCESS.2020.2973755, 2014.

[18]   F. Hdioud, B. Frikh and B. Ouhbi, "Multi-criteria Recommender Systems Based on Multi Attribute Decision Making," Proc. of the International Conference on Information Integration and Web-based Applications & Services (IIWAS'13), pp. 203-226, DOI:10.1145/2539150.2539176, 2013.

[19]   Y. Zheng, "Context-aware Collaborative Filtering Using Context Similarity: An Empirical Comparison," Information, vol. 13, no. 1, p. 42, DOI: 10.3390/info13010042, 2022.

[20]   Y. Shi, M. Larson and A. Hanjalic, "Mining Contextual Movie Similarity with Matrix Factorization for Context- aware Recommendation," ACM Transactions on Intelligent Systems and Technology (TIST), vol. 4, Article no. 16, pp. 1-19, DOI: 10.1145/2414425.2414441, 2013.

[21]   Y. Zheng and A. A. Jose, "Context-aware Recommendations *via* Sequential Predictions," Proc. of the 34th ACM/SIGAPP Symposium on Applied Computing (SAC'19), Limassol, Cyprus. pp. 2525-2528, DOI: 10.1145/3297280.3297639, 2019.

[22]   H. Al Tair, M. J. Zmerly, M. El Qutayry and M. Leida, "Architecture for Context-aware Pro-active Recommender System," Int. J. Multimedia and Image Proces. (IJIMP), vol. 2, no. 3/4, pp. 125-133, 2012.

[23]   F. Rezaeimehr et al., "TCARS: Time and Community-aware Recommender System," Future Generations Computer Systems, vol. 78, no. 1, pp. 419-429, 2017.

[24]   M.Unger, A. Bar, B. Shapira and L. Rokach, "Towards Latent Context-aware Recommendation Systems," Knowledge Based System, vol. 104, pp. 165-178, 2016.

[25]   L. Liu, N. Mehandjiev and L. Xu, "Using Contextual Information for Service Recommendation,"Proc. of the 44th Hawaii Int. Conference on System Sciences, pp. 1-9, DOI: 10.1109/HICSS.2011.476, 2011.

[26]   Y. Zheng, B. Mobasher and R. Burke, "Similarity-based Context-aware Recommendation," Proc. of the International Conference on Web Information Systems Engineering (WISE 2015), Part of the Lecture Notes in Computer Science Book Series, vol. 9418, DOI: 10.1007/978-3-319-26190-4_29, 2015.

[27]   E. Khazei and A. Alimohammadi, "Context-aware Group Oriented Location Recommendation in Location-based Social Networks," ISPRS International Journal of Geo-information, vol.8, no. 406, 2019.

[28]   A. V. Smirnov et al., "Group Context-aware Recommendation Systems," Scientific and Technical Information Processing, vol. 41, no. 5, pp 325-334, 2014.

[29]   T. M. Phuong, D. T. Lien and N. D. Phuong , "Graph- based Context-aware Collaborative Filtering," Expert Systems with Applications, vol. 126, pp. 9-19, DOI: 10.1016/j.eswa.2019.02.015, 2019.

[30]   H. X. Hiep et al., "Context-similarity Collaborative Filtering Recommendation," IEEE Access, vol. 8, pp. 33342-33351, DOI: 10.1109/ACCESS.2020.2973755, 2017.

[31]   S. Linda, S. Minz and K. K. Bharadwaj, "Effective Context-aware Recommendations Based on Context Weighting Using Genetic Algorithm and Alleviating Data Sparsity," Applied Artificial Intelligence, vol. 34, no. 10, pp.730-753, DOI: 10.1080/08839514.2020.1775011, 2020.

[32]   H. Ma, I. King and M. R. Lyu, "Effective Missing Data Prediction for Collaborative Filtering," Proc. of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp.39-46, 2007.

[33]   Y. Zheng, "Non-dominated Differential Context Modeling for Context-aware Recommendation," Applied Intelligence, vol. 2022, DOI: 10.1007/s10489-021-03027-5, 2020.

[34]   Y. Zheng, R. Burke and B. Mobasher, "Differential Context Relaxation for Context-aware Travel Recommendation," Proc. of the 13th International Conference on Electronic Commerce and Web Technologies (EC-WEB), vol. 123, pp. 88-99, 2012.

[35]   Github, "Movie_DePaulMovie," [Online], Available: https://github.com/irecsys/CARSKit/blob/master/context-aware_data_sets/Movie_DePaulMovie.zip.

[36]   Github, "Music_InCarMusic," [Online], Available:https://github.com/irecsys/CARSKit/blob/master/context-aware_data_sets/Music_InCarMusic.zip.

[37]   X. R. Garcia and G. M. Valdz, "Post-filtering for a Restaurant Context-aware Recommender System," Proc. of Recent Advances on Hybrid Approaches for Designing Intelligent Systems, Part of Studies in Computational Intelligence, vol. 547, pp. 695-707, 2014.

[38]   A. Kosir, A. Odic, M. Kunaver, M. Tkalcic and J.F. Tasic, "Database for Contextual Personalization," Elektrotehn Vestnik, vol. 78, no. 5, pp. 270-274, 2011.

[39]   B. Sunita, L. M. Aher and R. J. Lobo, "Data Mining in Educational System Using WEKA," Proc. of the International Conference on Education and Training Technologies (ICETT 2011), no. 3, pp. 20-25, 2011.

[40]   K. Saravananathan and T. Velmurugan, "Analysing Diabetic Data Using Classification Algorithms in Data Mining," Indian Journal of Science and Technology, vol. 9, no. 43, pp. 1-9, November 2016.

[41]   D. Chang, J. Liu, Z. Xu, H. Li, H. Zhu and X. Zhu, "Context-aware Tree-based Deep Model for Recommender Systems," Proc. of DLP-KDD, ACM NY, USA, DOI: 10.48550/arXiv.2109.10602, 2021.

[42]   B. Sunita, L.M. Aher and R.J. Lobo, "Comparative Study of Classification Algorithms," International Journal of Information Technology and Knowledge Management, vol.5, no. 2, pp. 239-243, 2012.

[43]   N. Bharagava et al., "Decision Tree Analysis on J48 Algorithm for Data Mining," International Journal of Advanced Research in Computer Science and Software Engineering, vol. 3, no. 6, 2013.

[44]   E. Ahishakiye, E.O Omulo, D. Taramwa and I. Niyonzima, "Crime Prediction Using Decision Tree (J48) Classification Algorithm," International Journal of Computer and Information Technology, vol.6, no. 3, pp. 188-195, 2017.

[45]   J. R. Quinlan, "Induction of Decision Tree," Machine Learning, vol. 1, pp. 81-106, 1986.

[46]   M. S. Tsechansky and F. Provost, "Handling Missing Values When Applying Classification Models," Journal of Machine Learning Research, vol. 8, pp. 1625-1657, 2007.

[47]   N. Q. Phan, P. H. Dang and H. X. Huynh, "Statistical Implicative Similarity Measures for User-based Collaborative Filtering Recommender System," Int. J. of Advanced Computer Science and Applications (IJACSA), vol. 7, no. 11, pp.140-146, 2016.

[48]   A. Gunawardana and G. Shani, "A Survey of Accuracy Evaluation Metrics of Recommendation Tasks," Journal of Machine Learning Research, vol. 10, pp.2935–2962, 2009.

[49]   J. L. Herlocker, J. A. Konstan et al., "Evaluating Collaborative Filtering Recommender Systems," ACM Transactions on Information Systems, vol. 22, no. 1, pp.5-53, 2004.

[50]   B. Ouhbi et al., "Deep Learning Based Recommender Systems," Proc. of the 5th IEEE International Congress on Information Science and Technology (CiSt), pp. 161-166, DOI: 10.1109/CIST.2018.8596492, Marrakech, Morocco, 2018.

[51]   S. Ahmadian, M. Ahmadian and M. Jalili, "A Deep Learning Based Trust-and Tag-aware Recommender System," Neurocomputing, vol. 488, pp. 557-571, DOI: 10.1016/j.neucom.2021.11.064, 2021.

ملخص البحث:

يُعدّ التنـائر مشـكلةً جدّيـةً في التّصـفية التعاونيـة التـي لهـا أثـر ملحـوظ فـي جـودة التّوصية. ويـتمّ إدخـال المعلومـات السّـياقية فـي أنظمـة التّوصـية التّقليديـة الـى جانـب معلومـات المسـتخدِمين ومعلومـات المفـرَدات للتّغلُّـب علــى هـذه المشـكلة. وقـد أثبـت العديـد مـن الأعمـال البحثيـة أنّ إدخـال المعلومـات السّـياقية قـد يزيـد مـن البيانـات المتنـاثرة. لـذا فـإنّ تقنيـات تعـدين البيانـات هـي مـن بـين الحلــول الفعّالـة التـي جـرى اسـتخدامُها فـي أنظمـة التّوصية الواعية للسّياق للتّعامل مع مشكلة تناثُر البيانات.

تقتــرح هــذه الورقــة دمْـج تقنيــة توصـية للتّصـفية التّعاونيـة تقـوم علــى تشـابُه السّـياقات والمسـتخدمين مـع تقنيـات تعـدين البيانـات كحـلٍّ لهـذه المشـكلة مـن أجـل تطـوير نظـام توصـية مبتكـر، وهـو نظـام مبنـي علـى القواعـد وواع للسّـياق. وقـد تمّـت تجربـة النّظـام المقتـرح بإدخـال أربـع خوارزميـات مبنيـة علـى القواعِـد علــى أربـع مجموعـات بياناتٍ مختلفـة، ومــن ثـمّ مقارنـة أداء النّظـام المقتـرح بـأداء النّمــاذج الــواردة فـي أدبيـات الموضـوع. وقـد بينـت نتـائج التّجربـة أنّ وزْن التّشـابُه مـع السّـياق ودمجـه مـع تقنيـة مبنيـة علـى القواعـد مـن الممكـن أن يقـودا الـى التّغلُّـب علـى مشـكلة تنـاثُر البيانـات ويعمـلا علـى تحسين دقّة التّوصية بالمقارنة مع النّماذج الواردة في أدبيات الموضوع.

المجلة الأردنية للحاسوب وتكنولوجيا المعلومات

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

عنوان المجلة

المجلة الأردنية للحاسوب وتكنولوجيا المعلومات

JJCIT