Princess Sumaya University for Technology
جامعة الأميرة سميّة للتكنولوجيا

صندوق دعم البحث العلمي والابتكار
Scientific Research and Innovation Support Fund

Jordanian Journal of Computers and Information Technology

December 2024 VOLUME 10 NUMBER 04
ISSN 2415 - 1076 (Online)
ISSN 2413 - 9351 (Print)

JJCIT

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

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

Jordanian Journal of Computers and Information Technology (JJCIT)

The Jordanian Journal of Computers and Information Technology (JJCIT) is an international journal that publishes original, high-quality and cutting edge research papers on all aspects and technologies in ICT fields.

JJCIT is hosted and published by Princess Sumaya University for Technology (PSUT) and supported by the Scientific Research Support Fund in Jordan. Researchers have the right to read, print, distribute, search, download, copy or link to the full text of articles. JJCIT permits reproduction as long as the source is acknowledged.

AIMS AND SCOPE

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

INDEXING

JJCIT is indexed in:

EDITORIAL BOARD SUPPORT TEAM

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

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 10, No. 04, December 2024.

INTERPRETING ARABIC TRANSFORMER MODELS: A STUDY ON XAI INTERPRETABILITY FOR QUR'ANIC SEMANTIC-SEARCH MODELS

Ahmad M. Mustafa, Saja Nakhleh, Rama Irsheidat and Raneem Alruosan

(Received: 10-Jan.-2024, Revised: 24-Mar.-2024, 16-May-2024 and 8-Jul.-2024, Accepted: 13-Jul.-2024)

ABSTRACT

Transformers have shown their effectiveness in various machine-learning tasks. However, their "black box" nature often obscures their decision-making processes, particularly in Arabic, posing a barrier to their broader adoption and trust. This study delves into the interpretability of three Arabic transformer models that have been fine-tuned for semantic-search tasks. Through a focused case study, we employ these models for retrieving information from the Holy Qur'an, leveraging Explainable AI (XAI) techniques—namely, LIME and SHAP—to shed light on the decision-making processes of these models. The paper underscores the unique challenges posed by the Qur'anic text and demonstrates how XAI can significantly boost the transparency and interpretability of semantic-search systems for such complex text. Our findings reveal that applying XAI techniques to Arabic transformer models for Qur'anic content not only demystifies the models' internal mechanics, but also makes the insights derived from them more accessible to a broader audience. This contribution is twofold: It enriches the field of XAI within the context of Arabic semantic search and illustrates the utility of these techniques in deepening our understanding of intricate religious documents. By providing this nuanced approach to the interpretability of Arabic transformer models in the domain of semantic search, our study underscores the potential of XAI to bridge the gap between advanced machine-learning technologies and the nuanced needs of users seeking to explore complex texts like the Holy Qur'an. Our code is available at[1].

KEYWORDS

Explainable machine learning, Semantic search, Arabic NLP, Transformers, SHAP, LIME.

1. INTRODUCTION

Interpretive and Explainable Artificial Intelligence (XAI), in the field of machine learning, deep learning and transformer models, has seen remarkable developments. However, XAI for Arabic transformer models remains a notable challenge. The idea of Explainable AI (XAI) emerged, introducing techniques that offer a reasonable trade-off between explainability and predictive power for a variety of machine-learning (ML), deep-learning and transformer techniques [1].

Transformer models, well known for their effectiveness in natural-language processing (NLP) tasks, often operate as black-box, making it difficult to understand the decision-making process they employ. This ambiguity sheds light on significant challenges, especially in the context of Arabic models, where small linguistic differences can increase the complexity of the interpretation task. The absence of robust XAI tools hinders the examination of model outputs, leading to a potential lack of trust and liability. Bridging this gap in interpretability for Arabic transformer models makes it possible for people to comprehend, trust and manage the newest generations of AI models in the Arabic-speaking world.

Arabic transformer models, used as black-box AI systems, have gained widespread usage in domains such as social networks, medicine and scientific fields. However, the necessity to explain and interpret these models arises from their operation as opaque decision making. These reasons include the Regulatory Perspective, exemplified by the European Union's General Data Protection Regulation (GDPR), which accords users the right to explanation. Another reason is the Model Developmental Perspective, which dives into issues such as limited training data, biased data, outliers, adversarial data and overfitting leading to inappropriate results in black-box AI systems. Lastly, the end-user and

[1] https://gist.github.com/a-mustafa/51fcacf30ecdf0c13ac91ad16fecfa89

A. Mustafa (Corresponding Author), S. Nakhleh, R. Irsheidat and R. Alruosan are with the Department of Computer Information Systems, Jordan University of Science and Technology, Irbid, Jordan. E-mails: ammustafa@just.edu.jo, swnakhleh21@cit.just.edu.jo, roirsheidat21@cit.just.edu.jo and rmalrosan21@cit.just.edu.jo

social perspectives address concerns about trust in black-box AI models, shedding light on the potential for unfair decisions and biases in the data used for model development. XAI is recognized as a solution to enhance trust by providing explanations, improving interpretability, addressing fairness concerns and ensuring that the models fulfill their intended purpose [2].

This study is motivated by the absence of XAI models for Arabic transformer models. As a case study, we perform a Qur'anic semantic search using different Arabic transformer models and then interpret them using different XAI models. The Holy Qur'an is the most significant source for Arabic and Islamic sciences. The Qur'an is considered a sacred text in Arabic and contains approximately 80,000 words divided into 114 chapters; each chapter consists of a varying number of verses. It also includes knowledge of a variety of other subjects, including science and the history of humanity [3]. Classical Arabic (CA), Modern Standard Arabic (MSA) and Colloquial Arabic are the three main styles or forms of the Arabic language [4]. Qur'an is the most important source of Classical Arabic. Many tools and applications have been developed to help in Qur'anic information retrieval.

There are three main methods for information retrieval within the Qur'an: semantic-based, keyword-based and Cross-Language Information Retrieval (CLIR) [5], as shown in Figure 1. A semantic-based method searches for concepts or meanings, whereas a keyword-based method looks for exact letter matches. CLIR searches for information in a language other than the one used in the query. Most Qur'an search tools use keyword search, but some use ontology-based or synonym-set methods [6]. The ontology-based or semantic-search approach looks for concepts or subjects that fit a user request. Semantic search emphasizes the meaning of words and the intent of the user query rather than relying only on keyword matching. It analyzes the context and considers the relationships between words and their meanings to retrieve similar information. Semantic search utilizes a transformer-based model such as BERT, neural models like RNN, ML models including n-gram and Word2Vec models [7].

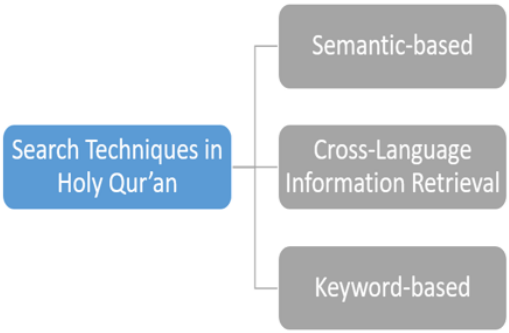


Figure 1. Classification of information-retrieval methods in the Holy Qur'an.

In this paper, we use three transformer BERT-based models for Qur'anic semantic search. Transformers are one of the most advanced techniques for many NLP problems since they were proposed by Vaswani et al. (2017) [8] for machine translation. The semantic-search models utilized in this study include: *CL-AraBERT* [7], an Arabic BERT transformer for CA. Additionally, "asafaya/bert-base-arabic" (*ArabicBERT*) developed by Safaya et al. [9] is employed. ArabicBERT is a pretrained language model based on BERT, designed for Arabic semantic-search task. Lastly, "multi-qa-MiniLM-L6-cos-v1" (*S-BERT*) model[2] [10] is used. S-BERT is a sentence-transformer model. It maps sentences and paragraphs to a 384-dimensional dense vector space and was designed for semantic search.

Although transformer-based neural networks excel at classification in various domains, they lack the capability to offer explanations for their predictions [11]. Our study shows different XAI techniques that interpret the transformers mentioned above using two SHAP [12] and LIME [13]. SHAP (SHapley Additive exPlanations) is an explainability technique that calculates the Shapley values from cooperative game theory to attribute the contributions of each feature to the model output, providing a comprehensive explanation for a given prediction. LIME (Local Interpretable Model-agnostic

[2] https://huggingface.co/sentence-transformers/multi-qa-MiniLM-L6-cos-v1

Explanations) is a model-agnostic technique that generates locally faithful approximations of a complex model decision boundaries by perturbing and observing input instances, facilitating interpretability for individual predictions. The main purpose of XAI is to introduce an explanation for a variety of ML, DL and transformer models that offer a reasonable trade-off between explainability and predictive power. This concept allows people to understand, trust and manage the newest generations of AI models.

1.1 Challenges of Qur'anic-text Processing

Challenges arise in the search and retrieval of relevant verses from the Holy Qur'an, due to both the search techniques and the structure of the text. The following are some of these challenges:

1. **Orthography**: The Qur'anic text employs a distinct orthography that incorporates the essential diacritics (tashkeel) and vowel marks (harakat) necessary to understand the text. Nevertheless, contemporary Arabic text frequently excludes these diacritics or stems them as part of preprocessing, posing challenges for machines to accurately recognize and handle the proper pronunciation. For example,الجَنّة (al-jannah) means heaven and الجِنّة (al-jinah) means ghosts [6]. The Qur'anic text utilizes a distinctive orthography (conventional spelling), distinct even from CA, referred to as al-rasm al-'Uthmani. This is the method of writing the Qur'anic text compiled during the reign of Caliph Uthman b. Affan, for example: أنزلناه (anzlnah \We revealed it), is written as أنزلنَه (anzlnah\We revealed it) [14].
2. **Textual Variants**: The Qur'anic text exists in various versions that can vary in spelling, pronunciation and significance. Consequently, developing reliable and uniform computational models for processing the Qur'anic text presents a significant challenge. For example, محمد (Muhammad), أحمد (Ahmad) and المُزّمّل (Mozzammil) all refer to Prophet Muhammad [15], as following: وَمُبَشِّرًا بِرَسُولٍ يَأْتِي مِن بَعْدِي اسْمُهُ أَحْمَدُ (wamubashshiran birasul yati min baedi asmuh 'ahmad \And bringing good tidings of a Messenger who will come after me, whose name will be Ahmad), مُّحَمَّدٌ رَّسُولُ اللَّهِ (Muhammad rasul alllah \Muhammad is the Messenger of Allah), يَا أَيُّهَا الْمُزَّمِّلُ (ya 'ayuha almuzzmmil \O you who wraps himself in clothing!)

   On the other hand, there exists a disparity between user inquiry, inscribed in MSA and retrieved Qur'anic verses, written in CA [16]. For example, searching for أنزلناه (anzlnah\We revealed it), in MSA should retrieve the word أنزلنَه (anzlnah\We revealed it) in CA. Since the vocabulary and spelling in CA differ from MSA, this makes models' selection challenging. To solve this issue, we select a multilingual S-BERT model, in addition to the CL-AraBERT model trained on CA and MSA texts, as mentioned before.
3. **Semantic Interpretation**: The Qur'anic scripture comprises numerous allegories, metaphors and parables that require deep semantic analysis and comprehension for accurate interpretation. For instance, the term الحيوان (Al-Hayawan\animal) in Arabic typically translates to 'the animal', but in this specific verse, it denotes 'the life'. و ما هذه الحياة الدنيا الا لهو و لعب و ان الدار الاخرة لهي الحيوان لو كانوا يعلمون (Wa ma hadhihi al-hayatu ad-dunya illa lahwan wa la'ab. Wa innad-dara al-akhirata lahiya al-hayawan law kanu ya'lamoon \And the worldly life is nothing but amusement and diversion. But the home of the Hereafter - that is the eternal life, if only they knew) [17]. To achieve a precise interpretation and translation of the Qur'anic text, it is imperative to understand its historical context. The text was revealed in the 7th century and the language and vocabulary used in it are indicative of the historical and cultural background of that period.
4. **Expressiveness**, which refers to rhetoric in linguistics, involves expressing meanings using fewer words. For example, the concise phrase فأسقيناكموه (Fa asqaynākumūhu), which translates to "and We have given it to you to drink" in Arabic morphology is remarkably intricate, yet follows a systematic approach [14].

Contribution: To the best of our knowledge, this study is the first to address the interpretation of Arabic semantic-search transformer models, utilizing *post-hoc* interpretation models. We propose a methodology that interprets the results obtained from three intricate transformer models; namely,

S-BERT, ArabicBERT and CL-AraBERT, which have recently been introduced for Qur'anic semantic search. Our results will help understand the inner workings of these Arabic semantic-search transformer models, facilitated by the utilization of *post-hoc* interpretation models, including LIME and two versions of SHAP, thus enhancing comprehension and insight.

2. RELATED WORK

This section provides an overview of previous studies that have explored semantic similarity in Arabic and Qur'anic text. It also delves into techniques for interpreting transformer models, such as BERT, utilizing *post-hoc* interpretation models such as SHAP and LIME. *Post-hoc* approaches refer to methods applied after a model has been trained to explain its predictions and provide insight into the decision-making process of complex models. These methods approximate the rationale of the underlying machine-learning models, proving especially valuable for the interpretation of 'black-box' models, wherein the internal mechanisms are not inherently transparent [18]-[19]. It is also worth mentioning that *ante-hoc* approaches, though not as widely recognized or discussed as *post-hoc* methods, *ante-hoc* approaches refer to techniques that are integrated during the model-development phase to ensure interpretability from the outset. These approaches are designed to build inherently explainable models, allowing for real-time interpretation of model decisions while processing data [20]. However, this approach is beyond the scope of our paper. Moreover, we evaluate and compare the *post-hoc* interpretation techniques with those used in our paper, highlighting their simplicity and informativeness.

Both topics discussed in this section are crucial to the research, as they provide the foundational knowledge and tools necessary for interpreting Arabic semantic-search transformer models, which are the main focus of this paper. Due to the lack of studies that combine Arabic semantic similarity with interpretation techniques, we have organized the related works into different sub-sections.

2.1 Semantic Similarity in Arabic and Qur'anic Texts

Several studies have employed different techniques to extract semantic similarity or relatedness from Arabic and Qur'anic texts. Alsaleh et al. (2021) [21] conducted experiments using the QurSim dataset and a fine-tuned AraBERT model, which is an Arabic-language model trained on a wide range of Arabic texts. The dataset includes pairs of verses classified into three classes: '2' for strong similarity, '1' for weak similarity and '0' for no similarity. They also filtered the dataset to eliminate repetition and create random pairs of verses. AraBERTv0.2 outperformed AraBERTv2 with an accuracy score of 92%. However, AraBERT struggled with classical-Arabic lexical synonyms and religious context, potentially due to corpus limitations. Our study utilizes AraBERT to classify pairs of Qur'anic verses as semantically related or not.

Mohamed and Shokry (2022) [6] discussed modern semantic-search techniques for the Holy Qur'an. They manually created a dataset and annotations based on Tajweed Mushaf and created an embedding matrix trained with classical Qur'anic and Arabic texts. This generated word-based feature vectors for the verses. During queries, cosine similarity was used to find the most semantically similar result. However, this approach only retrieved verses for the first query and ignored the rest of the topics, although they are also relevant to the query.

Saeed et al. (2020) [22] explored using word embeddings to identify semantically similar verses from the Holy Qur'an. Using Word2Vec and Sent2Vec models, they highlighted the importance of semantic text similarity in NLP and various fields, including religious-text analysis. They trained custom word embeddings from multiple English translations of the Holy Qur'an and compared them to pre-trained embeddings from the Spacy library. The custom-trained models showed promising performance, with Model #5 achieving the highest accuracy. The study emphasized the framework's potential to be applied to any text, contributing to a deeper understanding of sacred and literary works. Notably, their research focused on English translations of the Holy Qur'an, potentially missing nuances in the original Arabic.

Malhas and Elsayed (2022) [7] proposed the first Qur'anic Reading Comprehension Dataset (QRCD), consisting of 1,337 question-passage-answer triplets for 1,093 question-passage pairs. They introduced CLassical-AraBERT (CL-AraBERT), pre-trained on a 1.0B-word classical Arabic dataset to

complement modern standard Arabic (MSA) resources, enhancing its utility for reading comprehension tasks. Leveraging cross-lingual transfer learning from MSA to classical Arabic, they fine-tuned CL-AraBERT using MSA-based machine-reading comprehension datasets followed by QRCD. For evaluation, they used the F1-score and Partial Average Precision (pAP), integrating partial matching for multi-answer and single-answer MSA questions, thus constituting the first MRC system on the Holy Qur'an.

2.2 Interpretation Techniques

Although there are many studies related to Qur'anic semantic search, there is no previous work that interprets Arabic semantic-search models using XAI techniques. Several *post-hoc* XAI interpretation techniques are discussed here to interpret and explain different transformer models.

The first technique is LIME [13], which generates local explanations for each instance in a dataset. LIME introduces disturbances to an instance and uses the newly generated dataset to predict the class of each instance using a trained classifier. A simpler model is then used to explain the classifier's prediction. While LIME is likely to be locally faithful, it does not perfectly represent complex models.

SHAP [23], another *post-hoc* XAI technique, interprets the complex behavior of machine-learning, deep-learning and transformer models. SHAP values, based on game theory, allocate importance scores to each feature within a model to provide consistent explanations. Positive SHAP values indicate a positive contribution to the prediction, whereas negative values indicate a negative impact. The Semantic Textual Similarity (STS) explainer [24] is a SHAP-based technique designed to explain sentence-level scores by highlighting erroneous words in both source and target sentences. This method helps understand the contribution of each word using SHAP for tasks like machine translation and semantic search involving different text languages. TransSHAP, proposed by Kokalj et al. (2016) [11], adapts SHAP to provide sequential explanations for transformer models such as BERT-based text classifiers. Unfortunately, it is notable that TransSHAP is currently not compatible with semantic-search transformer models. Despite not being compatible with semantic-search transformer models, TransSHAP was found effective for tasks like sentiment analysis. It was rated better than SHAP and slightly better than LIME in overall user preferences.

Layer-wise Relevance Propagation (LRP) [1] assigns relevance scores to input features to explain machine-learning model predictions. When applied to Transformer models, LRP computes relevance scores for each input token to understand its contribution to the final prediction. Although useful, LRP, like TransSHAP, faces limitations in providing explanations for tasks involving multiple-sentence analysis, such as semantic search.

El Zini et al. (2022) [25] proposed new metrics and techniques to evaluate the explainability of Arabic Sentiment Analysis (SA) models. They assessed the accuracy of 'rationales' extracted by the model and compared the agreement between XAI techniques and human judgment on a dataset. Their results showed that transformer models have better explainability than convolutional and recurrent neural-network architectures. This research lays the foundation for designing interpretable NLP models and creating a common evaluation framework.

3. METHODOLOGY

3.1 Dataset

The Holy Qur'an, revered by 1.5 billion Muslims worldwide, is structured into 30 sections and 114 chapters, encompassing 6,236 verses, totaling approximately 78,000 words. These words are organized into verses, with sets forming parts, chapters and groups (Hizb) or Hizb quarters. Each of the 114 chapters belongs to one of the 30 sections and the text is further segmented into 60 groups (Hizb), with each section comprising two groups (Hizb) [7].

We have used a verified Qur'an dataset called Tanzil Quran text[3]. The Tanzil Quran text provides a verified digital version of the Holy Qur'an in many scripting styles, including the Uthmani style. We have utilized the normalized simple-clean text style (in Tanzil 1.0.2) to enable the use of the dataset with transformer-based language models that have already been pre-trained using normalized Arabic

[3] https://tanzil.net/download/

text. Tanzil Qur'an dataset consists of three columns, as shown in Table 1: 1. Surah ID: is an id for each Surah from 1 to 114. 2. Verse ID: is an id for each verse (ayah) from 1 to 6236 without verse Basmallah, except in Chapter 1 (Surah Al-Fatiha). 3. Verse Text: the content of verse text with diacritics. For the model evaluation using the Qur'an exegesis step, we used an official Qur'an exegesis (Tafsir) called QuranEnc[4]. Qur'anEnc is a dataset that provides an interpretation for each verse of the Qur'an. As shown in Table 2, there are three columns in Qur'an exegesis (QuranEnc) dataset: 1. Verse ID: is an id for each verse (ayah) from 1 to 6236. 2. Exegesis: the content of verse's exegesis (tafsir). 3. Verse Text: the content of verse text with diacritics.

Table 1.The-holy-Qur'an dataset.

Surah ID	Verse ID	Verse Text
1	1	بِسْمِ اللَّهِ الرَّحْمَـٰنِ الرَّحِيمِ (Bismillāhi al-Raḥmāni al-Raḥīm\In the name of Allah, the Most Gracious, the Most Merciful)
1	2	الْحَمْدُ لِلَّهِ رَبِّ الْعَالَمِينَ (Al-ḥamdu lillāhi rabbi al-'ālamīn\Praise be to Allah, the Lord of all the worlds)
1	3	الرَّحْمَـٰنِ الرَّحِيمِ (Ar-Raḥmāni ar-Raḥīm\The Entirely Merciful, the Especially Merciful)

Table 2. Qur'an exegesis (QuranEnc) dataset.

Verse ID	Exegesis	Verse Text
3122	وإن ربك, أيها الرسول, لهو العزيز الذي ينتقم من أعدائه، الرحيم بمن تاب من عباده (Wa inna rabbaka, ayyuha ar-rasulu, lahu al-'azizu alladhi yantaqimu min 'a'ada'ihi, ar-rahīmu biman tāba min 'ibādihi\And indeed, your Lord, O Messenger, He is the Exalted in Might, the One who exacts retribution upon His enemies, yet He is the Merciful to those among His servants who repent and mend their ways)	وَإِنَّ رَبَّكَ لَهُوَ الْعَزِيزُ الرَّحِيمُ (Wa innna rabbaka lahuwa al-'azīzu ar-raḥīmu.\And indeed, your Lord is the Exalted in Might, the Merciful)
4465	في بساتين وعيون جارية (Fī basātīn wa 'uyūn jāriyah \In gardens and flowing springs)	فِي جَنَّاتٍ وَعُيُونٍ (Fi jannatin wa 'uyun \In gardens and springs)
5888	واستمعت لربها منقادة، وحق لها ذلك (Wa istam'at li rabbiha munqādah, wa ḥaqqun lahā dhālik \And she listened to her Lord obediently. It was rightful for her to do so)	وَأَذِنَتْ لِرَبِّهَا وَحُقَّتْ (Wa 'adhīnat li rabbiha wa ḥuqqat \And she listened to her Lord and fulfilled [her obligation])

3.2 Data Preprocessing

The lack of diacritics in Modern Standard Arabic (MSA) is a common issue in the Arabic language. Diacritical marks are significant, because they impact the meaning and subsequently, the comprehension of Arabic texts [26]. Although the Holy Qur'an is extensively diacritical, most NLP tasks involving digital Qur'anic text resort to normalization by eliminating diacritics during the preparation stage. In this phase, we applied several preprocessing techniques to clean the text before feeding it into models using the Holy Qur'an dataset. Firstly, we added a new column named "surah name" to the dataset, which includes the name of each surah in the Holy Qur'an. Following this addition, we removed tashkeel (diacritical marks) and tatweel (character lengthening), as well as eliminating stop words and punctuation from the verses. Lastly, we normalized certain characters to standardize the dataset. Table 3 illustrates examples of the data preprocessing steps.

Table 3. Overview of data preprocessing steps with examples.

Original Verses	بِسْمِ اللَّهِ الرَّحْمَـٰنِ الرَّحِيمِ (Bismillāhi al-Raḥmāni al-Raḥīm\In the name of Allah, the Most Gracious, the Most Merciful)	فَوَاكِهُ ۖ وَهُمْ مُكْرَمُونَ (Fawākihu, wahum mukramūn \Fruits and they are honored)
Tashkeel removing	بسم الله الرحمن الرحيم	فواكه وهم مكرمون
Tatweel removing	بسم الله الرحمن الرحيم	-
Punctuation and stop words removing	-	فواكه مكرمون

[4] https://quranenc.com/ar/browse/arabic_mokhtasar/

3.3 Workflow

As discussed earlier, the aim of this study is to interpret Arabic BERT-based semantic-search models using LIME and two SHAP techniques, the most well-known XAI techniques. Observe Figure 2.

Semantic search is designed to understand the meaning of a user query, as opposed to simply matching keywords and to return results that are relevant to the user intent. This can make search results more accurate and useful to the user. Semantic-search technology is used in a variety of applications, including search engines, e-commerce websites and voice assistants. We will evaluate the semantic-search models using two different methods, BERTScore and Cosine similarity and then interpret the model results through the XAI techniques.

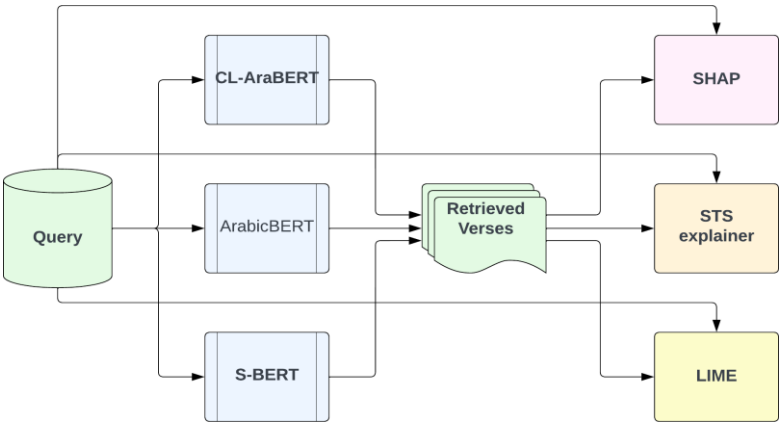


Figure 2. Proposed framework: A query is sent to semantic-search models and the retrieved verses are interpreted alongside the query using XAI techniques.

For the Qur'anic semantic search, first, we passed all 6,236 Qur'an exegesis, as queries to the three transformer BERT-based models: CL-AraBERT, ArabicBERT and S-BERT. The verses retrieved by the three models have been compared with the reference verse recorded in the Qur'anEnc Tafsir dataset. In this step, BERTScore precision, recall and F1-score were calculated for all the verses and their exegesis. BERTScore metric [27] evaluates the quality of text embeddings, particularly in the context of comparing the generated text against reference text. Specifically, it compares token-level similarity and leverages contextual embeddings from BERT or other transformer-based models.

The performance of each model has been evaluated using BERTScore precision, recall and F1-score measurements. BERTScore is an automated evaluation metric that is used to assess the quality of text-generation systems. The precision (P), expressed in Equation1, measures the mean cosine similarities between each retrieved token and its closest reference token, normalized by the number of retrieved tokens. Using contextual embeddings, tokens are represented in a reference verse $x = x_1, \ldots, x_k$ and a retrieved verse $\hat{x} = \hat{x}_1, \ldots, \hat{x}_l$. The cosine similarity $(x_i^T \hat{x}_j)$ weighs each retrieved token. Recall (R) indicates the extent of coverage completeness, as shown in Equation 2, calculated by dividing the number of relevant retrieved tokens by the number of all possible related tokens. The F1-score is the harmonic mean of precision and recall as shown in Equation 3.

$$P = \frac{1}{|\hat{x}|} \sum_{\hat{x}_j \in \hat{x}} \max_{x_i \in x} (x_i^T \hat{x}_j) \tag{1}$$

$$R = \frac{1}{|x|} \sum_{x_i \in x} \max_{\hat{x}_j \in \hat{x}} (x_i^T \hat{x}_j) \tag{2}$$

$$F = \frac{2 * P * R}{P + R} \tag{3}$$

Cosine similarity [28] measures the similarity between two nonzero vectors in an inner product space. In NLP, it is commonly employed to evaluate the similarity between two pieces of text by converting each text into a vector of word counts or frequencies and finding the cosine of the angle between the vectors. The cosine similarity helps identify the degree of alignment of these vectors, indicating semantic similarity. This allows for efficient retrieval of sentences (verses) with similar meanings, making it a valuable metric for tasks such as semantic search.

In the second evaluation, samples of exact, similar and dissimilar queries were sent to these three semantic-search models and the results were evaluated based on the cosine similarity between each query result and reference verse.

Finally, sample queries, along with their respective search results, underwent three interpretation techniques: SHAP, STS explainer and LIME. Two criteria were employed in selecting the *post-hoc* interpretation techniques. Firstly, the XAI technique should support the Arabic transformer models. Secondly, it should support tasks that involve comparing two sentences, such as machine translation, question answering and semantic search. Notably, both TransSHAP and LRP were excluded from consideration, since they do not support tasks that involve comparing two sentences. The interpretation step will explain the two evaluation methods mentioned previously for semantic-search models. This process involves three steps: First, a query is sent to each semantic-search model. Then, three results are chosen (exact, similar and dissimilar) based on their scores. Finally, each resulting verse is compared with the query using the interpretation techniques. More details will be explained in the next sections.

4. EXPERIMENTS

In this section, we outline the experimental setup utilized in our study, focusing specifically on the methodology adopted for interpreting the semantic-search model outcomes and reporting the ultimate findings. To test the three semantic-search models, we have developed our testing procedure, where we use the sentences mentioned in an official Qur'an exegesis (Tafsir), Qur'anEnc. We fed all 6,236 verses of interpretation texts (Tafsir) into the semantic-search models. If the reference verse mentioned in Tafsir is retrieved among the closest five resulting verses, we consider it as the prediction; otherwise, we consider the top retrieved verse as the prediction. Subsequently, we compare these predictions with the references from the Tafsir dataset. This procedure is repeated for all semantic-search models.

4.1 Models Evaluation Using Exact, Similar and Dissimilar Sample Queries

Since BERT is a transformer-based model, its embeddings are contextual and depend on the entire input sequence, so that SHAP can be adapted to work with BERT models by approximating the Shapley values for token embeddings. Therefore, calculating Shapley values directly becomes computationally expensive. To handle this limitation, we have employed sampling sub-sets of input tokens to estimate Shapley values for the three BERT models. To do so, we have passed a set of 3 queries (samples) to three models: CL-AraBERT, ArabicBERT and S-BERT. Then, we measured the cosine similarity between the query and the retrieved verses. After that, we interpret samples of them, using SHAP and STS-Explainer interpretation techniques in sub-section 4.2. Here, as a first experiment, we have applied 3 test cases to validate each of the 3 BERT models:

1. We have passed an existing text such as علمه البيان (’Alamahu al-bayan \He taught him eloquence), as the expected results are the exact match with a similarity of 1.0. The other retrievals should be other similar sentences, but not identical. As expected, their cosine scores will be less than the exact match.
2. We have passed a text that does not exist in the Qur'an, but similar to existing words such as ابراهام (Ibrāhīm\Abraham). Here, the expected results should be verses with words similar to the query.
3. The third query uses words neither exist nor are similar to Qur'anic words such as كمبيوتر (Kumbiutir \Computer).

4.2 Model Interpretation Using SHAP

For each query, all semantic-search transformer models typically operate by retrieving a set of results, prioritizing the exact matches if they exist, followed by similar results and then possibly dissimilar ones. In this step, we interpret 3 samples from the retrieved verses for exact match, similar and dissimilar results using *post-hoc* interpretation techniques, SHAP and STS explainer.

First, we search for an existing sentence, such as: علمه البيان (’Alamahu al-bayan \He taught him eloquence), so that we could select 3 results:

1. Exact match with cosine similarity scores 1.0, such as: علمه البيان (’Alamahu al-bayan \He taught him eloquence).
2. Similar verse with high cosine similarity score, such as علم القران (’Allama al-Qur’an \He taught the Quran).
3. A verse with a very low cosine similarity score, such as إنما جزاء الذين يحاربون الله ورسوله ويسعون في الأرض فساداً أن يقتلوا (Innama jazau alladhina yuharibuna Allah wa rasulahu wa yas’awna fi al-ardi fasadan an yuqtalu \Indeed, the penalty for those who wage war against Allah and His Messenger and strive upon earth [to cause] corruption is none but that they be killed).

All these results were picked for the 3 BERT models. So in total, we have 9 results that were passed to 2 SHAP interpretation techniques. For the first SHAP model, we have tuned a Question Answering SHAP technique, to work as a semantic-search model. We assumed that the question was our query علمه البيان (’Alamahu al-bayan \He taught him eloquence) and sent it with the context, which was the Surah or verse that contained the query. Finally, we have assigned scores to the selected three search results to display their relevance.

SHAP explains the output of the semantic search by attributing the importance of each feature in the context to the model prediction (query result). The SHAP summary plots provide many visual details that thoroughly explain the machine-learning models in a simple way. First, the summary plots provide SHAP scores, f(input), which equals the summation of all the word (feature) scores in the context. They provide insights into the contribution of each feature in the context to the model prediction for a particular instance of data, in our case, the query result. The words in the context, with a positive SHAP score suggest a positive influence on the prediction (query result), while the negative score suggests a negative influence on the prediction. The magnitude of SHAP score provides a measure of the feature importance relative to other features in the input query. Features with higher SHAP values are considered more important in influencing the model prediction. Second, SHAP colors in the summary plots are important to investigate the SHAP scores. The red color means a positive effect, the blue color means a negative effect and the shade of the color indicates the amount of effect. Therefore, dark red means a high positive effect, while light red means a low positive effect.

In the "STS explanation" step, we interpret the results from another perspective, where another technique was utilized from SHAP called STS Explainer. We passed the same three search results from the "Search" step to the STS Explainer which was implemented especially for semantic-search tasks. The similarity score metric of the STS Explainer is F1 by default and we have fixed it for all the upcoming experiments. f(x) shows the similarity between the query and the result, while E(f(X)) shows the Expected SHAP score, which is calculated as the mean of all predictions. Just like the original SHAP technique, the STS Explainer provides visual interpretation using colors to indicate the positive or negative impact of the values, red for positive and blue for negative. STS Explainer also provides the SHAP score for each word in both the query and the result sentences, which indicates each word contribution to the model prediction that is figured out in the summary plot.

4.3 Interpretation Using LIME

In this study, we utilized a surrogate model, specifically LIME, to interpret the outputs generated by BERT models when analyzing the Holy Qur’an. We examined two variations of the Qur’anic text: the original text, which includes tatweel (elongation marks), tashkeel (diacritical marks), punctuation and stop words and a second version where only the tashkeel was removed (the Tanzil Qur’an dataset).

The Holy Qur’an dataset underwent a detailed normalization process, as described in sub-section 3.2, which includes steps such as stop-word removal and character normalization. The text (verses) resulting from this process is referred to as "Normalized Verses" in Table 4. In contrast, the Tanzil Qur’an version underwent a simpler process, with only the removal of diacritical marks (tashkeel). The text resulting from this process is labeled as "Normalized Verses (Tashkeel Removed)" in Table 4. The primary motivation for these different approaches was to investigate the impact of text normalization on model interpretation and similarity assessment. This methodological choice allowed us to directly compare how varying levels of text normalization influence the performance and interpretability of BERT models.

To facilitate this comparison, we encoded the verses from these two variations using the Sentence

Transformer of our transformer models (CL-AraBERT, ArabicBERT and S-BERT) to get the embedding. The Sentence Transformer is a deep-learning model that encodes text into high-dimensional vector representations (embeddings) to capture their semantic meaning, facilitating efficient comparison and analysis of text data[5]. These embeddings were then used to calculate cosine similarity with the verse علمه البيان (’Alamahu al-bayan \He taught him eloquence), serving as a benchmark for assessing verse similarity.

We adopted a binary classification approach to present these similarities, designating verses as "*Similar*" (label ’1’) or "*Not Similar*" (label ’0’) based on predefined thresholds of 0.6, 0.8 and 1.0[6]. This classification facilitated a structured analysis of the impact of text normalization at varying levels of strictness in similarity assessment. The verses were then split into training and testing datasets. The outcomes of this approach are detailed in Table 4, which presents the classification results (labels) of selected normalized and normalized (tashkeel removed) verses at different similarity thresholds.

Table 4. Classification outcomes of selected verses at varied similarity thresholds.

Verses	Similarity threshold	Similarity value (label)
Normalized Verses		
بسم الله الرحمن الرحيم - علمه البيان Bismillāhi ar-Raḥmāni ar-Raḥīm \In the name of Allah, the Most Gracious, the Most Merciful	>=0.8	0
الحمد لله ربّ العالمين - علمه البيان Al-ḥamdu lillāhi rabbi al-’ālamīn\Praise be to Allah, the Lord of all the worlds	>=0.8	0
فألقى عصاه فإذا هي ثعبان مبين - علمه البيان Fa’alqā ’aṣāhu fa’idhā hiya thu’ban mubīn \So he threw down his staff and behold! it was a manifest serpent	>=0.8	1
الذين جعلوا القران عضين - علمه البيان Alladhīna ja’alu al-qur’āna ’idīn\Those who have made the Quran burdensome	>=0.8	1
بسم الله الرحمن الرحيم - علمه البيان Bismillāhi ar-Raḥmāni ar-Raḥīm \In the name of Allah, the Most Gracious, the Most Merciful	>=0.6	0
الحمد لله ربّ العالمين - علمه البيان Al-ḥamdu lillāhi rabbi al-’ālamīn\Praise be to Allah, the Lord of all the worlds	>=0.6	0
مالك يوم الدين - علمه البيان Māliki yawmi ad-dīn \Master of the Day of Judgment	>=0.6	1
ذلك الكتاب لا ريب فيه هدى للمتقين - علمه البيان Dhālika al-kitābu lā rayba fīhi hudan lil-muttaqīn \This is the Book about which there is no doubt, a guidance for those conscious of Allah	>=0.6	1
علمه البيان - علمه البيان ’Allamahu al-bayan \He taught him eloquence	==1	1
Normalized Verses (Tashkeel Removed)		
بسم الله الرحمن الرحيم - علمه البيان Bismillāhi ar-Raḥmāni ar-Raḥīm \In the name of Allah, the Most Gracious, the Most Merciful	>=0.8	0
الحمد لله رب العالمين - علمه البيان Al-ḥamdu lillāhi rabbi al-’ālamīn\Praise be to Allah, the Lord of all the worlds	>=0.8	0
وإن عليك اللعنة إلى يوم الدين - علمه البيان Wa inna ’alayka al-la’natu ilá yawmi ad-dīn \And upon you is the curse until the Day of Recompense	>=0.8	1
وأن عذابي هو العذاب الأليم - علمه البيان Wa anna ’adhābī huwa al-’adhābu al-aleem \And indeed, My punishment is the painful punishment	>=0.8	1
بسم الله الرحمن الرحيم - علمه البيان Bismillāhi ar-Raḥmāni ar-Raḥīm \In the name of Allah, the Most Gracious, the Most Merciful	>=0.6	0
الذين يؤمنون بالغيب ويقيمون الصلاة ومما رزقناهم ينفقون - علمه البيان Alladhīna yu’minūna bil-ghaybi wa yuqīmūna aṣ-ṣalāh wa mimma razaqnāhum yunfiqūn \Those who believe in the unseen, establish prayer and spend out of what We have provided for them	>=0.6	0
الحمد لله رب العالمين - علمه البيان Al-ḥamdu lillāhi rabbi al-’ālamīn\Praise be to Allah, the Lord of all the worlds	>=0.6	1
مالك يوم الدين - علمه البيان Māliki yawmi ad-dīn \Master of the Day of Judgment	>=0.6	1
علمه البيان - علمه البيان ’Allamahu al-bayan \He taught him eloquence	==1	1

[5] https://sbert.net/

[6] If the similarity between verses and the verse (query) (’Alamahu al-bayan \He taught him eloquence) is greater than or equal to the similarity threshold, we set the similarity class to “Similar" else “Not Similar". For example, suppose that we set the similarity threshold to 0.8 and the similarity score is 0.7, the similarity value will be “Not Similar" since 0.7 is less than the threshold.

In Arabic, two words that appear different on the surface can become similar when represented using word embeddings due to the language's rich morphological structure. Arabic is characterized by its introflexive, fusional and inflectional nature, which means that words can have different surface forms, but share underlying roots and patterns that convey similar meanings or functions [29]-[30]. For instance, in verses such as فَأَلْقَىٰ عَصَاهُ فَإِذَا هِيَ ثُعْبَانٌ مُّبِينٌ (So he threw down his staff and behold, it was a manifest serpent) and عَلَّمَهُ الْبَيَانَ (He taught him eloquence), the words البيان (al-bayan) and مبين (mubīn) look different but share the same root ن –ي– ب (b-y-n), which conveys the meaning of clarity, explanation or distinction. When these words are embedded in a vector space, the embeddings capture these morphological and semantic similarities, making their vectors close to each other. This closeness can be quantified using cosine similarity.

Our analysis involved a Logistic Regression (LR) model to predict the similarity class (label) of each verse in the test dataset. This model was chosen for its simplicity and effectiveness in binary-classification problems. Specifically, we trained the LR model on the training dataset and utilized TF-IDF (Term Frequency-Inverse Document Frequency) [31] to convert the textual data into numerical vectors. Finally, the predicted probabilities from the LR model were then passed to the LIME XAI framework for interpretability and explanation of the model's decisions.

5. RESULTS AND DISCUSSION

We delve into the outcomes of the conducted experiments. To understand the intent and context of a user query, we pass two evaluation techniques to evaluate both semantic-search models retrieving the meaning and the matching keywords. First, the model evaluation using Qur'an exegesis using BERTScore and second, the model evaluation using exact, similar and dissimilar sample queries using cosine similarity.

5.1 Model Evaluation Using Qur'an Exegesis

As mentioned in the experimental-setup section, we compared the three semantic-search model predictions with the references from the Tafsir dataset. This procedure is repeated for all semantic-search models. Table 5 shows that CL-AraBERT outperformed the other two models with 0.92, 0.93 and 0.92 for Precision, Recall and F1 BERTScore measurements, respectively.

Table 5. Test scores for 3 different semantic-search models.

	CL-AraBERT	ArabicBERT	S-BERT
P-$_{\text{BERTScore}}$ Eq. 1	0.92	0.79	0.67
R-$_{\text{BERTScore}}$ Eq. 2	0.93	0.81	0.71
F1-$_{\text{BERTScore}}$ Eq. 3	0.92	0.80	0.69

5.2 Model Evaluation Using Exact, Similar and Dissimilar Sample Queries

We have passed a set of three sample queries -exact match, similar and dissimilar- to three models: CL-AraBERT, ArabicBERT and S-BERT. Then, the cosine similarity is measured between the query and the retrieved verses. The results for these experiments are shown in Table 6.

The exact-match results show identical similarity for the three models, with a cosine score of 1.0 or approximately 1 (0.99) for all of them. Even for the other retrieved results, as will be shown in experiment 2, they were close to each other. For the similar word, the expected retrieval from a linguistic perspective , is: إبراهيم (Ibrāhīm \Abraham). However, the retrieved verses contain similar sub-words (or tokens), not the same word, that got similar embeddings, such as token: "را" (ra). Their cosine scores were not very high, which indicates that they are far from the query embedding. For the third query, since we have searched for a non-existent word, the retrieved sentences contained similar tokens. Hence, the retrieved result for S-BERT, for example, contained sub-token: "بيـ" (byo) . The proposed results through this experiment have obviously shown that the BERT transformers got the same cosine similarity scores, for the exact-match results. Also, variations have been shown for the other queries due to their different embedding, even if all of their implementations were BERT-based.

Table 6. The top retrieved verse for each query using the three models along with cosine similarity.

BERT model	علمه البيان: exact match ('Alamahu al-bayan \He taught him eloquence)	إبراهام: similar (Ibrāhīm\Abraham)	كمبيوتر: not exist (Kumbiutir\Computer)
S-BERT	علمه البيان	فالمدابرات أمرا (Fāl-dabirāt amrā\So the consequences are decreed.)	كراما كاتبين (Karaman kātibīn\Honored recorders)
	0.99	0.83	0.93
ArabicBERT	علمه البيان	إلا المصلين (Illā al-muṣallīn\Except the ones who pray)	عسق ('Ayn, Seen, Qaf\letters, none but Allah (Alone) knows their meanings)
	1	0.87	0.84
CL-AraBERT	علمه البيان	إله الناس (Ilāh al-Nās\God of the people)	فيها كتب قيمة (Fīhā kutubun qīmah\In it (are) valuable books)
	1	0.76	0.7

5.3 Model Interpretation Using SHAP

In this technique, we explain the functioning of the model output. For the CL-AraBERT model, after retrieving the exact match and hovering over it among the three options, the resulting SHAP score was 0.03. This score represents the summation of all scores attributed to the contextual features (words). Referring to Figure 3, the blue words in the context negatively impacted the score of the علمه البيان ('Alamahu al-bayan \He taught him eloquence) result, while the red tokens positively influenced it.Specifically, red words force the SHAP value towards the positive side (arrow from left to right), while blue words push the SHAP values from right to left, towards the negative side. The negative and positive values shown correspond to the line, representing the magnitude of the influence exerted by each word. To compare the similar result with other results of lower and upper similarities, observe Table 7.

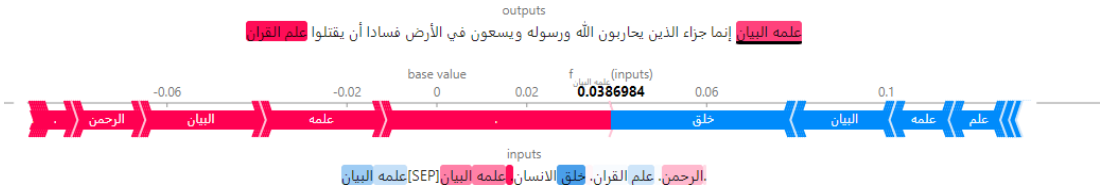


Figure 3. SHAP values for exact-match retrieval for CL-AraBERT model.

After hovering over the similar result علم القران ('Allama al-Qur'an\He taught the Quran), as shown in Figure 4, the obtained SHAP score was 0.04. This score was determined by summing the SHAP scores for both the red and blue words, which, respectively, influenced the positive and negative SHAP scores. To compare the similar result with other results of lower and upper similarities, observe Table 7.

Figure 4. SHAP values of the query similar retrieval for CL-AraBERT model.

In cases where dissimilar results were observed, it was noted that the SHAP score is 0. This suggests that the SHAP values of the words in the context were very low, less than 0.002, indicating their negligible impact on the prediction of the SHAP score for the dissimilar results. Figure 5 illustrates the

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 10, No. 04, December 2024.

SHAP score of the dissimilar retrievals.

إنما جزاء الذين يحاربون الله ورسوله ويسعون في الأرض فساداً أن يقتلوا

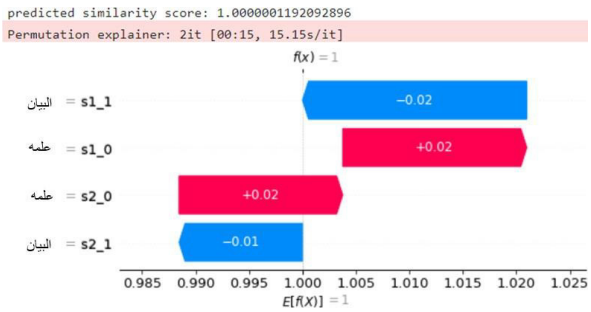
(Innama jazau alladhina yuharibuna Allah wa rasulahu wa yas'awna fi al-ardi fasadan an yuqtalu \ Indeed, the penalty for those who wage war against Allah and His Messenger and strive upon earth [to cause] corruption is none but that they be killed)

علمه البيان

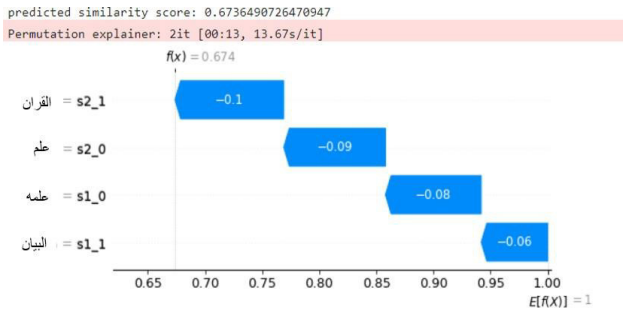
('Alamahu al-bayan \He taught him eloquence)

Figure 5. SHAP values of the query dissimilar retrieval for CL-AraBERT model.

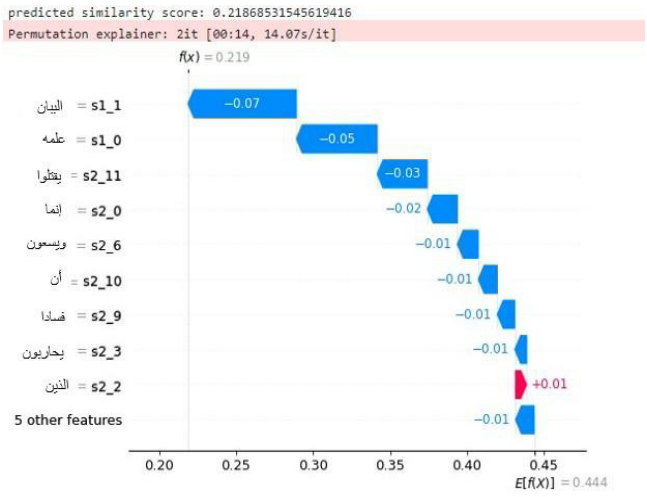
These outcomes show obviously that the SHAP explanation has fitted the semantic-search outcomes in addition to the visual interpretation for each vector effect. All details for the 3 models are mentioned in the Supporting Materials [7]. Despite these results, tuning SHAP QA technique revealed some shortcomings. For example, the word علمه ('Alamahu\He taught him) gave both positive and negative indications simultaneously in calculating the SHAP score for similar results, as shown in Figure 5. This can be attributed to the intricate nature of language and the contextual nuances present. As SHAP utilizes BERT model to score and interpret results, it is expected to reveal words with varying impacts simultaneously. BERT score tokens are based on the surrounding context and the overall tone of the text, which can result in certain words having mixed effects. Additionally, the SHAP score of the exact-match obtained values lower than those of similar results, which contrasts with the outcome of semantic-search evaluation either using the Qur'anic exegesis or using the cosine similarity.



(a) Exact match



(b) Similar retrieval



(c) Non-convergence

Figure 6. STS Explainer scores for CL-AraBERT model.

[7] https://github.com/SajaNakhleh/Quranic-semantic-search

Consequently, in the "STS explanation" step, we passed the same three search results from the "Search" step to the STS Explainer. For the exact-match results, all STS Explainer scores matched the predicted similarity, even if sometimes with a very minor difference (less than 0.01). Figure 6a shows the same scores for similar words with the same effect on the predicted score, which is the value of f(x). The expected scores E(f(x)) were close to the predicted similarity scores f(x), as shown in Fig. 6b. However, the presence of the phrase علم القران ('Allama al-Qur'an\He taught the Quran) in a sentence resulted in a lower score of 0.67 and negatively affected the model prediction. For the nonconvergence results, expected score E(f(x)) was far from the predicted score f(x) and SHAP scores, even if the nonconvergence tokens got high negative affect on the predicted scores. Observe Figure 6c.

Table 7 summarizes the results of 3 scores: cosine similarity score from "Search" step, SHAP scores from "SHAP Interpretation" step and STS Explainer similarity scores from "STS Explainer Interpretation" step. The observed results strongly support the notion that the outcomes of Arabic BERT model are consistent with SHAP results. While the positive SHAP scores (i.e., +0.04 and +0.03) indicate a direct relationship; while the neutral SHAP score indicates dissimilar results. STS Explainer has presented results that align with the cosine-similarity results of the Arabic BERT model. The relationship is direct and positive, as observed from the table (i.e., 0.67 for similar results and 0.21 for dissimilar results).

Table 7. Scores of SHAP interpretation and STS explanation experiments for CL-AraBERT.

ID	Result	Cosine Sim.	QA SHAP	STS Sim.
A	علمه البيان 'Alamahu al-bayan \He taught him eloquence	1	+0.03	1
B	علم القرآن 'Allama al-Qur'an\He taught the Qur'an	0.71	+0.04	0.67
C	إنما جزاء الذين يحاربون الله ورسوله و يسعون في الأرض فساداً أن يقتلوا Innama jazau alladhina yuharibuna Allah wa rasulahu wa yas'awna fi al-ardi fasadan an yuqtalu. The penalty for those who wage war against Allah and His Messenger and strive upon earth to cause corruption is none but that they be killed	0.12	Neutral	0.21

5.4 Interpretation Using LIME

In this sub-section, we utilize the LIME framework to interpret the predictions made by our logistic regression model in the S-BERT model. Specifically, LIME helps explain why the model classified certain verses as similar (label '1') or not similar (label '0'). The insights gained from this interpretative step are summarized in Table 8. This table presents a comparative analysis of the S-BERT model's performance across various similarity thresholds, taking into account the effects of normalization. The model's assessment is conducted on a single verse from the Holy Quran, showcasing its normalized form, the applied similarity threshold, the predicted probability, which indicates the likelihood assigned by the model to a particular class for the given instance and the resulting similarity classification.

Preliminary findings highlight a nuanced impact of text normalization on similarity assessment. Particularly, the verses evaluated at a similarity threshold of >= 0.8 in Table 8 did not exhibit discernible differences between normalized and normalized (tashkeel removed) versions with a similarity result of 0. This can be attributed to the high similarity threshold and the significant impact of each word (feature) in predicting the "*Not Similar*" class. For instance, in the verse يدريك يزكي, which has been normalized, both words contributed to the "*Not Similar*" classification. Similarly, in the verse وما يدريك لعله يزكى, which has been tashkeel removed, the words يدريك, لعله and يزكى had the highest impact on predicting the "*Not Similar*" class.

In contrast, a slight divergence emerged among verses assessed at a similarity threshold of >= 0.6. For instance, in the verse يدريك يزكي, which has been normalized, the word يزكى contributed to the "*Not Similar*" class, achieving a prediction probability of 0.89, while the word يدريك contributed to the "*Similar*" class, achieving a prediction probability of 0.11. Similarly, in the verse وما يدريك لعله يزكى,

which has been tashkeel-removed, the words وما, يدريك and يزكى had the highest impact on prediction as "*Not Similar*", achieving a prediction probability of 0.79, whereas the word لعله contributed to "*Similar*" class, achieving a prediction probability of 0.21.

Interestingly, the analysis also indicated that text normalization significantly impacts the assessment of text similarity in Arabic. The process of normalization, which includes more than one technique such as tashkeel removing, tatweel removing and punctuation and stop-word removal, has been shown to improve determining similarity of text by ensuring that only the most distinctive lexical features are retained.

These observations underscore the complex nature of text normalization's influence on semantic analysis when employing BERT models. The findings suggest that while normalization can facilitate the identification of superficial textual similarities, it might obscure deeper semantic relationships present in the unaltered text. This insight opens up new avenues for research, particularly in the development of more nuanced-text preprocessing techniques that balance the need for normalization with the preservation of semantic richness. Future studies could explore alternative approaches to text preprocessing and their effects on model interpretability and performance, further enriching our understanding of the intricate dynamics between text normalization and NLP models.

Table 8. Comparative analysis of S-BERT model performance across various similarity thresholds considering normalization effects. Values in **bold** represent the prediction probability for class *'Not Similar* (0)', while values in parentheses indicate the prediction probability for class '*Similar* (1).'

Models	Verses	Normalized Forms	Similarity threshold	Prediction Probability	Similarity Result
S-BERT	وَمَا يُدْرِيكَ لَعَلَّهُ يَزَّكَّى Wa mā yudrikal-la'allahu yazzakkā \And what can make you know? Perhaps he [will] purify himself	يدريك يزكي	>=0.8	**1.00** (0.00)	0
	وَمَا يُدْرِيكَ لَعَلَّهُ يَزَّكَّى Wa mā yudrikal-la'allahu yazzakkā \And what can make you know? Perhaps he [will] purify himself	يدريك يزكي	>=0.6	**0.89** (0.11)	0
	وَمَا يُدْرِيكَ لَعَلَّهُ يَزَّكَّى Wa mā yudrikal-la'allahu yazzakkā \And what can make you know? Perhaps he [will] purify himself	وما يدريك لعله يزكى	>=0.8	**1.00** (0.00)	0
	وَمَا يُدْرِيكَ لَعَلَّهُ يَزَّكَّى Wa mā yudrikal-la'allahu yazzakkā \And what can make you know? Perhaps he [will] purify himself	وما يدريك لعله يزكى	>=0.6	**0.79** (0.21)	0

In our experiments, we have compared the performance of our proposed explainability technique on a semantic-search task with the previously proposed explainability techniques by El Zini et al. [25] on a similar task. Our method has used only AraBERT-based models, while the method by El Zini et al. used eight different models for English texts, two of which were BERT-based. We have found that our method achieved very close outcomes. El Zini et al. used an additional XAI technique called: anchor, while we have used STS Explainer as an additional XAI technique.

The results are different from sentiment analysis, which is a classification task. Our task measures the scores for the query *versus* the retrieved results. Furthermore, the accuracy of the model is not measured in the sentiment-analysis tasks proposed by El Zini et al. Our results demonstrate that SHAP and LIME align with the BERT transformers for the Arabic language. However, it is important to note that our experiment is limited to a specific domain (i.e., Qur'anic text) and further research is needed to generalize the findings.

6. Conclusion and Future Work

We study the opacity of Arabic transformer models using SHAP and LIME interpretation techniques, applying these to benchmark Qur'anic semantic search within the Qur'anEnc dataset. Our findings reveal that SHAP interpretations align closely with BERT model predictions, highlighting their effectiveness in predicting correct results. Specifically, our experiments demonstrated that SHAP and STS Explainer scores correlate with cosine similarity in exact-match retrievals, with CL-AraBERT showing significant positive effects for exact matches. Interestingly, nonconvergence retrievals exhibited divergent scores, suggesting areas for further investigation. Additionally, our analysis using

LIME of text normalization impact on BERT models' performance revealed that unnormalized texts yield more logical similarity scores in certain instances. These insights not only shed light on the interpretability of Arabic transformer models, but also underscore the nuanced influence of text normalization on semantic-search tasks. In the future, our objective is to extend our examination to include AraT5 and AraGPT models, thereby enhancing our understanding and interpretation of Arabic transformers. This endeavor will undoubtedly contribute to the robustness and reliability of future Arabic-language processing tools.

REFERENCES

[1] A. Ali, T. Schnake, O. Eberle, G. Montavon, K.-R. Müller and L. Wolf, "XAI for Transformers: Better Explanations through Conservative Propagation," Proc. of the 39th Int. Conf. on Machine Learning, ser. Proc. of Machine Learning Research, vol. 162, pp. 435–451, PMLR, [Online], Available: https://proceedings.mlr.press/v162/ali22a.html, 17–23 Jul 2022.

[2] W. Saeed and C. Omlin, "Explainable AI (XAI): A Systematic Meta-survey of Current Challenges and Future Opportunities," Knowledge-based Systems, vol. 263, p. 110273, 2023.

[3] H. U. Khan, S. M. Saqlain, M. Shoaib and M. Sher, "Ontology Based Semantic Search in Holy Quran," International Journal of Future Computer and Communication, vol. 2, no. 6, p. 570, 2013.

[4] I. Al-Huri et al., "Arabic Language: Historic and Sociolinguistic Characteristics," English Literature and Language Review, vol. 1, no. 4, pp. 28–36, 2015.

[5] M. Mustafa, H. AbdAlla and H. Suleman, "Current Approaches in Arabic IR: A Survey," Proc. of the Int. Conf. on Asian Digital Libraries, Digital Libraries: Universal and Ubiquitous Access to Information, Part of the Book Series: Lecture Notes in Comp. Science, vol. 5362, pp. 406–407, 2008.

[6] E. H. Mohamed and E. M. Shokry, "QSST: A Quranic Semantic Search Tool Based on Word Embedding," J. of King Saud Uni.-Computer and Inform. Sciences, vol. 34, no. 3, pp. 934–945, 2022.

[7] R. Malhas and T. Elsayed, "Arabic Machine Reading Comprehension on the Holy Qur'an Using Clarabert," Information Processing and Management, vol. 59, no. 6, p. 103068, DOI:10.1016/j.ipm.2022.103068, 2022.

[8] A. Vaswani et al., "Attention Is All You Need," arXiv: 1706.03762, DOI: 10.48550/arXiv.1706.03762, 2023.

[9] A. Safaya, M. Abdullatif and D. Yuret, "KUISAIL at SemEval-2020 Task 12: BERT-CNN for Offensive Speech Identification in Social Media," Proc. of the 14th Workshop on Semantic Evaluation, Barcelona (online): Int. Committee for Computational Linguistics, pp. 2054–2059, Dec. 2020.

[10] N. Reimers and I. Gurevych, "Making Monolingual Sentence Embeddings Multilingual Using Knowledge Distillation," Proc. of the 2020 Conf. on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, arXiv: 2004.09813, 2020.

[11] E. Kokalj, B. Škrlj, N. Lavrač, S. Pollak and M. Robnik-Šikonja, "BERT Meets Shapley: Extending SHAP Explanations to Transformer-based Classifiers," Proc. of the EACL Hackashop on News Media Content Analysis and Automated Report Generation, Association for Computational Linguistics, pp. 16–21, [Online]. Available: https://aclanthology.org/2021.hackashop-1.3, Apr. 2021.

[12] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," Proc. of the 31st Int. Conf. on Neural Information Processing Systems (NIPS'17), pp. 4768–4777, Red Hook, USA, 2017.

[13] M. T. Ribeiro, S. Singh and C. Guestrin, ""Why Should I Trust You?": Explaining the Predictions of Any Classifier," Proc. of the 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD '16), pp. 1135–1144, New York, NY, USA, DOI: 10.1145/2939672.2939778, 2016.

[14] M. H. Bashir et al., "Arabic Natural Language Processing for Qur'anic Research: A Systematic Review," Artificial Intelligence Review, vol. 56, no. 7, pp. 6801–6854, 2022.

[15] A. Farghaly and K. Shaalan, "Arabic Natural Language Processing: Challenges and Solutions," ACM Trans. on Asian Language Inform. Processing, vol. 8, no. 4, DOI: 10.1145/1644879.1644881, 2009.

[16] K. Dukes and N. Habash, "Morphological Annotation of Quranic Arabic," Proc. of the 7th Int. Conf. on Language Resources and Evaluation (LREC'10), Valletta, Malta, ELRA, May 2010.

[17] S. Altammami and E. Atwell, "Challenging the Transformer-based Models with a Classical Arabic Dataset: Quran and Hadith," Proceedings of the 13th Language Resources and Evaluation Conf., pp. 1462–1471, European Language Resources Association, Marseille, France, Jun. 2022.

[18] D. Vale, A. El-Sharif and M. Ali, "Explainable Artificial Intelligence (XAI) Post-hoc Explainability Methods: Risks and Limitations in Non-discrimination Law," AI and Ethics, vol. 2, no. 4, pp. 815–826, 2022.

[19] E. M. Kenny, E. D. Delaney, D. Greene and M. T. Keane, "Post-hoc Explanation Options for XAI in Deep Learning: The Insight Centre for Data Analytics Perspective," Proc. of Int. Conf. on Pattern Recognition, ICPR Int. Workshops and Challenges, Part of the Book Series: Lecture Notes in

Computer Science, vol. 12663, pp. 20–34, 2021.
[20] A. Sarkar, D. Vijaykeerthy, A. Sarkar and V. N. Balasubramanian, "A Framework for Learning Ante-hoc Explainable Models *via* Concepts," Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 10 286–10 295, June 2022.
[21] A. Alsaleh, E. Atwell and A. Altahhan, "Quranic Verses Semantic Relatedness Using AraBERT," Proc. of the 6th Arabic Natural Language Processing Workshop, pp. 185–190, Kyiv, Ukraine, [Online]. Available: https://aclanthology.org/2021.wanlp-1.19, Apr. 2021.
[22] S. Saeed, S. Haider and Q. Rajput, "On Finding Similar Verses from the Holy Quran Using Word Embeddings," Proc. of the 2020 IEEE Int. Conf. on Emerging Trends in Smart Technologies (ICETST), pp. 1–6, Karachi, Pakistan, 2020.
[23] S. M. Lundberg et al., "Explainable Machine-learning Predictions for the Prevention of Hypoxaemia During Surgery," Nature Biomedical Engineering, vol. 2, no. 10, p. 749, 2018.
[24] C. Leiter, P. Lertvittayakumjorn, M. Fomicheva, W. Zhao, Y. Gao and S. Eger, "Towards Explainable Evaluation Metrics for Machine Translation," Journal of Machine Learning Research, vol. 25, pp. 1-49, 2023.
[25] J. El Zini, M. Mansour, B. Mousi and M. Awad, "On the Evaluation of the Plausibility and Faithfulness of Sentiment Analysis Explanations," Proc. of the IFIP Int. Conf. on Artificial Intelligence Applications and Innovations, Part of the Book Series: IFIP Advances in Information and Communication Technology, vol. 647, pp. 338–349, 2022.
[26] N. Habash, Introduction to Arabic Natural Language Processing, 1st Edn., ser. Synthesis Lectures on Human Language Technologies, Morgan and Claypool Publishers, vol. 3, pp. 1-124, 2010.
[27] T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger and Y. Artzi, "BERTscore: Evaluating Text Generation with BERT," Towards Data Science, arXiv: 1904.09675, 2020.
[28] Y. Qiao, C. Xiong, Z. Liu and Z. Liu, "Understanding the Behaviors of BERT in Ranking," arXiv preprint, arXiv: 1904.07531, 2019.
[29] S. Yagi, A. Elnagar and S. Fareh, "A Benchmark for Evaluating Arabic Word Embedding Models," Natural Language Engineering, vol. 29, no. 4, p. 978–1003, 2023.
[30] B. Dahy, M. Farouk and K. Fathy, "Arabic Sentences Semantic Similarity Based on Word Embedding," Proc. of the 2022 20th Int. Conf. on Language Engineering (ESOLEC), vol. 20, pp. 35–40, Valencia, Spain, 2022.
[31] G. Salton, "Introduction to Modern Information Retrieval," ISBN 0-07-054484-0, McGraw-Hill, 1983.

ملخص البحث:

تغوص هذه الدّراسة في تفسير ثلاثة نماذج تحويل باللّغة العربية تمّ تطويرها وتكييفها للمهامّ المتعلّقة بالبحث في علم دلالات الألفاظ. فمن خلال دراسة حالةٍ مركّزة، نقوم بتوظيف تلك النّماذج في استرجاع معلوماتٍ من القرآن الكريم، باستخدام تقنيتي (LIME) و (SHAP)، لإلقاء الضّوء على عمليات اتّخاذ القرار في النّماذج المدروسة. وتؤكّد الدّراسة التّحدّيات الفريدة التي تفرضها النّصوص العربية، وبخاصّةٍ القرآنية، كما تُبين أنّ النّماذج المدروسة تعمل على زيادة الشّفافية وقابلية التّفسير لأنظمة البحث في علم دلالات الألفاظ، وبخاصّةٍ بالنّسبة للنّصّ القرآني. وقد أثبتت النّتائج أنّ استخدام التقنيات المذكورة من شأنه أن يعمل على توضيح الآليات الدّاخلية للنّماذج قيد البحث، بالإضافة إلى جعل الاستنتاجات المستخلصة منها متاحةً لشريحةٍ أوسع من الجمهور. ويمكن القول إنّ مساهمة هذه الدّراسة في مجال البحث هي ذات وجهين. فمن جهة، تُغني المجال فيما يتعلق بالبحث في علم دلالات الألفاظ في النّصوص باللّغة العربية، وبخاصّة النّصّ القرآني، ومن جهة أخرى فهي تبيّن الفائدة من التّقنيات المستخدمة في تعزيز فهمنا للنّصوص والوثائق الدّينية.

والجدير بالذّكر أنّ هذه الدراسة من شأنها أن تُسهم في جَسْر الهُوّة بين تقنيات تعلُّم الآلة المتقدمة واحتياجات المستخدمين الّذين يسْعَون إلى استكشاف نصوصٍ "معقّدة"، كما هو النّصّ القرآني.

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 10, No. 04, December 2024.

LOCAL FEATURE SELECTION USING THE WRAPPER APPROACH FOR FACIAL-EXPRESSION RECOGNITION

Fatima Zohra Boukhobza[1], Abdenour Hacine Gharbi[2], Khaled Rouabah[3] and Phillipe Ravier[4]

(Received: 15-Apr.-2024, Revised: 29-Jun.-2024, Accepted: 16-Jul.-2024)

ABSTRACT

Automatic Facial Expression Recognition (FER) systems provide an important way to express and interpret the emotional and mental states of human beings. These FER systems transform the facial image into a set of features to train a classifier capable of distinguishing between different classes of emotions. However, the problem often posed is that the extracted feature vectors sometimes contain irrelevant or redundant features, which decreases the accuracy of the induced classifier and increases the computation time. To overcome this problem, dimensionality must be reduced by selecting only the most relevant features. In this paper, we study the impact of adding the "Wrapper" selection approach and using the information provided by different local regions of the face such as the mouth, eyes and eyebrows, on the performance of a traditional FER system based on a local geometric feature-extraction method. The objective here is to test and analyze how this combination can improve the overall performance of the original traditional system. The obtained results, based on the Multimedia Understanding Group (MUG) database, showed that the FER system combined with the proposed feature-selection strategy gives better classification results than the original system for all four classification models; namely, K-Nearest Neighbor (KNN) classifier, Tree classifier, NB classifier and Linear Discriminant Analysis (LDA). Indeed, a considerable reduction (up to 50%) in the number of features used and an accuracy of 100%, using the LDA classifier, were observed, which represents a significant improvement in terms of computation time, efficiency and memory space. Furthermore, the majority of relevant features used are part of the "eyebrows' region", which proves the importance of using information from local regions of the face in emotion recognition tasks.

KEYWORDS

Facial expression recognition, Local feature extraction, Feature selection, Wrapper approach, Classification.

1. INTRODUCTION

In the field of FER, which is one of the most important topics in image processing and computer vision, feature selection is a crucial step used to extract the most relevant information from images, thus improving performance of the model in various modalities, including two-dimensional (2D), three-dimensional (3D) and temporal data. For 2D FER, feature selection involves identifying and choosing relevant facial features that capture the information needed for expression recognition while minimizing irrelevant data. Whereas, for complex and diverse data, such as 3D and temporal facial expressions, where different angles and movements must be considered, a specific feature-selection approach tailored to the unique feature of each data type, allows the development of more precise and more robust models having a positive effect on the overall performance of the FER [1]-[2] system.

In the context of the 2D FER, several techniques for feature selection have been proposed in the literature. For example, in [3] the authors used the Relief-F feature-selection method to select the top-ranked features for each facial expression. In this method, features are classified in sequential order based on their variance values. After that, the selected features are used to train and test six classifiers which are Multi-Layer Perceptron (MLP), Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM), K-Nearest Neighbor (KNN) and Radial Basis Function (RBS). The results given in [3] showed that the KNN classifier is the most accurate classifier among the others with a total accuracy rating equal to 94.93% using the Cohn-Kanade (CK+) database.

Mahmood [4] applied both the Chi-square and Relief-F feature-selection methods to select the top-

1. F. Boukhobza is with LIST Laboratory, M'Hamed Bougara University, Boumerdes, Algeria. Email: f.boukhoubza@univ-boumerdes.dz
2. Hacine Gharbi is with Department of Electronics, MSE Laboratory, Mohamed El-Bachir El-Ibrahimi University, El-Anasser, Algeria. Email: abdenour.hacinegharbi@univ-bba.dz
3. K. Rouabah is with Department of Electronics, ETA Laboratory, Mohamed BOUDIAF University, M'sila, Algeria. Email: khaled.rouabah@univ-msila.dz
4. P. Ravier is with PRISME Laboratory, Université d'Orléans, France. Email: philippe.ravier@univ-orleans.fr

ranked features. The latter ones were then used to distinguish the most accurate classifier among the four classifiers which are the SVM, the KNN, the DT and the RBS. The obtained experimental results, based on the CK+ database, showed that KNN is the most accurate classifier for both feature-selection methods. Indeed, the average accuracy rate, corresponding to the KNN classifier, is 94.18% for the Chi- square method and 94.93% for the Relief-F method.

Ewees et al. [5] proposed an optimization method, called Sine-Cosine Algorithm (SCA) for feature selection. This method, which is a metaheuristic algorithm that utilizes the mathematical forms of sine and cosine functions to perform the optimization problems, was used to improve the performance of a real-time system intended to recognize student's facial expression. The performance of this method was evaluated on a database of Students' Facial Expression (SFE), using various classifiers including KNN, RF and SVM. The experimental results showed that the KNN classifier achieved the best performance with a remarkable reduction in the number of features.

Cossetin et al. [6] proposed a hybrid method to select relevant features and reduce dimensionality as well as computational complexity. In this method, the authors combined the "Filter" approach, in which features were ranked by relevance using techniques such as Correlation-based Feature Selection (CFS), Relief-F and Information Gain (IG), with the "Wrapper" approach. The latter one makes it possible to progressively optimize the sub-set of the most discriminating features for each pair of expressions. The results obtained in [6] demonstrated that this combination overcomes the limitations of existing methods while avoiding reliance on reference points and improving SVM classification accuracy.

Dino and Abdulrazzaq [7] compared different feature-selection methods for FER using multi-classification algorithms such as KNN, MLP, DT (J48) and NB. The comparison here, the main objective is to determine the best classifier capable of achieving an acceptable accuracy rate, is carried out on the basis of three feature-selection techniques; namely, CFS, the Gain Ratio (GR) and IG. The experimental results, based on the CK+ database, showed that the KNN classifier was the best, with an accuracy of 91% using only 30 features.

Wang et al. [8] proposed a feature-fusion model of multiple feature selection, realized by the measure of the RV correlation coefficient, to increase the performance of the FER system. Four kinds of features; namely, Local Binary Patterns (LBP), Pyramid of Histogram of Oriented Gradients (PHOG), Histogram of Oriented Gradients (HOG) and Gabor features are selected for fusion. In addition, Canonical Correlation Analysis (CCA) and Principal Component Analysis (PCA) sub-spaces are used to fuse selected features. The experimental results, based on the Japanese Female Facial Expression (JAFFE) and CK databases (using an SVM classifier) demonstrated that the proposed model improves performance and eliminates feature redundancy.

Li et al. [9] proposed an Iterative Universum Twin SVM (IUTWSVM) using Newton method for the classification of multiclass emotions by means of facial-expression analysis. This innovative approach, which integrates an iterative method, aims to improve feature selection and classification performance while reducing computational costs. To improve Universum's data selection and ensure that it effectively contributes to the classification task, the authors introduced a new system based on information entropy. Experimental results based on benchmark databases demonstrated that the IUTWSVM approach outperforms existing algorithms, in terms of classification accuracy and learning efficiency. This is true for both binary and multiclass facial emotion-recognition tasks.

Talaat et al. [10] developed a real-time emotion-identification system to detect autism spectrum disorders (ASDs) in autistic children. The proposed system integrates an autoencoder-based feature-selection method implemented in a Deep Convolutional Neural Network (DCNN) architecture. The results obtained demonstrated that this autoencoder achieved a remarkable accuracy of 95.23% with the Xception model. Additionally, it improves real-time emotion detection, thus providing an advanced technological solution adaptable on various smart devices.

Always in the context of improving the performance of FER systems, the authors in [11] used Active Shape Models (ASM) proposed by Cootes et al. in [12] to study the positioning of 24 facial landmarks (four marks for the mouth, four marks for each eye and six marks for each eyebrow). At this stage, several classifiers were also used to determine which of them gave the best results in identifying seven emotions; namely, happiness, anger, sadness, surprise, disgust, fear and neutral. The results presented

in [11] demonstrated that this approach has advantages in terms of robustness and interpretability. Additionally, it is stable to illumination changes and is invariant to global transformations such as face rotation and scaling.

To sum up, each of the various methods has its own set of limitations. For instance, Relief-F is a supervised method that calculates feature weights based on their relevance to other features, but it may struggle with noisy data. Chi-square measures the relationship between features and the target variable through statistical analysis. Nevertheless, it is less effective for non-linear relationships. The SCA employs mathematical functions for feature selection and global optimization, but can converge to local optima. Splitting a multi-class problem into binary ones simplifies the classification process, but may overlook the relationships between classes. Correlation evaluates the linear relationship between features and the target, but may not capture non-linear associations. GR and IG, which rely on entropy, can detect non-linear relationships, but might favor features with more values. RV correlation is a non-parametric method that captures both linear and non-linear relationships, although it can be computationally intensive. The autoencoder approach, based on neural networks, excels at learning non- linear feature relationships, but requires substantial data and computational resources [13]-[16]. Finally, for ASM-based approach, using the set of all extracted features, without selecting the relevant ones, increases the complexity in terms of computation time and memory space and likely reduces the accuracy caused by the curse of dimensionality.

The aim of this work is to demonstrate the efficiency and impact of adding a feature-selection step to the FER process, on the one hand and to conduct, on the other hand, a study on the effect of local facial regions such as the mouth, eyes and eyebrows on the performance of the FER system. This study will make it possible to determine the local facial regions containing the most important local features to ensure good recognition of facial emotional expressions. To achieve this, in this paper, we propose a modified enhanced FER system based on that given in [11]. In other words, we will integrate a "Wrapper" approach to remove redundant or irrelevant features that can degrade the classification model's performance and processing time. The reminder of this paper is organized as follows. Section 2 presents the related work. Section 3 describes the proposed modified enhanced FER system. In section 4, the experimental results, based on the MUG database, together with a performance comparison are given. We end with a conclusion.

2. RELATED WORK

A schematic overview of the proposed Silva et al. FER [11] system is shown in Figure (1). As illustrated in this figure, the latter system includes three functional blocks. The first one characterizes the face-detection function and the second is facial feature extraction and normalization. Finally, the last functional block represents facial-expression classification.

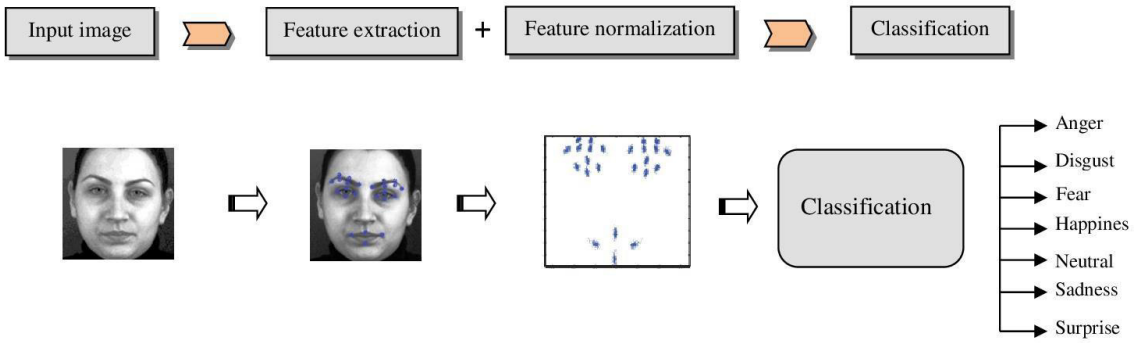


Figure 1. A schematic overview of the traditional FER system proposed in [11].

The principle of each of these blocks is given below.

2.1 Face-detection Block

In [11], faces are located using the algorithm proposed by Viola and Jones [17], which consists of two main steps: Haar-like feature extraction and AdaBoost classification.

2.2 Feature-extraction Block

After the face-detection step, features are extracted using the ASM method proposed by Cootes et al. [12]. ASM is a feature-matching method based on a statistical model [18] comprising two stages. In the first one, each structured object is represented by a set of landmarks manually placed in each image. In the second stage, the landmarks are automatically aligned to minimize the distance between their corresponding points. Then, the ASM creates a statistical model of the facial shape that iteratively deforms to fit the model in a new image [11]. The optimal number of facial landmarks ensuring better facial expression-recognition performance is chosen 24 [11]. This number, which ensures the maximum precision ratio, was chosen after an evaluation test based on 9 sets of facial landmarks containing 4, 12, 20, 24, 27, 35, 39, 55 and 76 landmarks. Table 1 describes the number of landmark points extracted for each facial region utilized.

Table 1. Number of landmark points extracted for each facial region.

Facial region	Number of landmark points
Mouth	4
Left eye	4
Right eye	4
Left eyebrow	6
Right eyebrow	6
Total	**24**

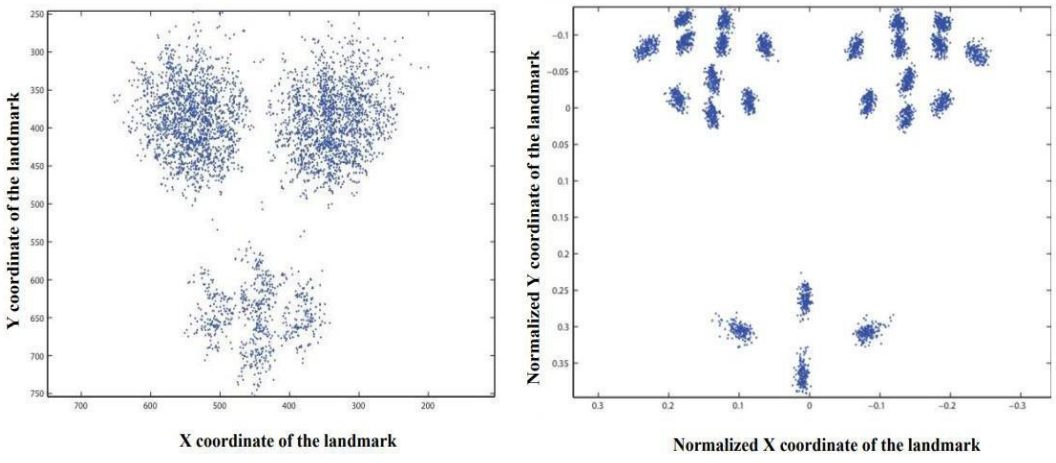
The choice of facial regions is based on the fact that facial expressions are deformations of the main permanent areas of the face (mouth, eyes, …etc.). These local areas contain most of the emotional information [19]-[20], which makes it possible to clearly recognize different emotions according to different deformations.

2.3 Feature-normalization Block

In this step, the Generalized Procrustes Analysis (GPA) is used to accomplish the following:

- Normalizing the extracted landmarks;
- Eliminating the effects of scale, rotation and translation in the landmarks' set;
- Decreasing the variations between the corresponding landmarks and search for the best fit.

GPA is a standard multivariate statistical method widely applied in shape analysis. To find the optimal superposition of two or more configurations, GPA involves three transformations; namely, translation, rotation and scaling [21]. Figure (2) shows the features extracted from the surprise expression using the MUG database. Figure 2(a) shows the raw data and Figure 2(b) shows the data normalized with GPA.



(a) (b)

Figure 2. Landmark normalization: (a) Raw data (b) Normalization with GPA [11].

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 10, No. 04, December 2024.

2.4 Facial-expression Classification Block

The final step of this FER system is the classification of facial expressions, which is achieved in [11] using several classifiers; namely, KNN [22], NB [23], LDA [24] and DT [15] classifiers. First, the KNN algorithm, one of the most popular algorithms for pattern recognition, predicts the test sample's category according to the K training samples that are its nearest neighbors, then determines the category by selecting the one with the highest category probability [25]. Second, NB is a statistical classifier based on Bayes' theorem, assigning the most likely class to a given example described by its feature vector. Third, LDA involves finding the projection hyperplane that minimizes the interclass variance and maximizes the distance between the projected means of the classes [26]. Finally, the DT classification algorithm establishes the relationship between the input attributes and the output attributes to build a model that predicts the desired class with the highest accuracy [27].

3. PRINCIPLE OF THE PROPOSED ENHANCED FER SYSTEM

Figure (3) shows the proposed modified schematic diagram of the FER system. As illustrated in this figure, our main contribution is the adoption of a feature-selection phase for the original FER system to determine how much its overall performance can be improved.

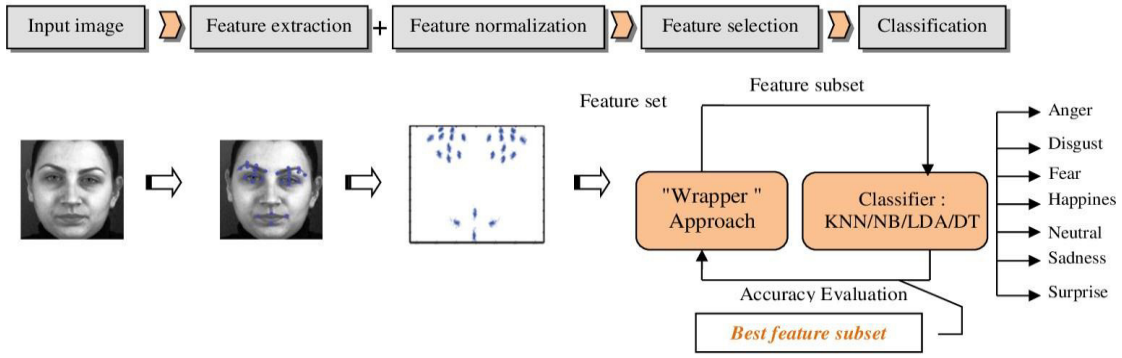


Figure 3. Schematic overview of our proposed modified FER system.

According to this FER system, using both ASM and GPA methods, 48 features (48 = 24 × 2: representing the Cartesian coordinates of 24 landmark points) are extracted from the face image and are also normalized. The feature vector is then constructed according to the following equation:

$$F_{\overrightarrow{i,j}} = (f_1, f_2, \ldots\ldots, f_n) \quad (1)$$

Where:
i and j denote the i-th landmark set for the j-th facial expression; $n = 24$, is the number of extracted landmarks; $f = (x, y)$, is the Cartesian coordinates for landmark f.

The 48 local features, resulting from the feature extraction, are given as follows:

$$m_{x1}, m_{y1}, m_{x2}, m_{y2}, m_{x3}, m_{y3}, m_{x4}, m_{y4}$$
$$le_{x1}, le_{y1}, le_{x2}, le_{y2}, le_{x3}, le_{y3}, le_{x4}, le_{y4}$$
$$re_{x1}, re_{y1}, re_{x2}, re_{y2}, re_{x3}, re_{y3}, re_{x4}, re_{y4}$$
$$lw_{x1}, lw_{y1}, lw_{x2}, lw_{y2}, lw_{x3}, lw_{y3}, lw_{x4}, lw_{y4}, lw_{x5}, lw_{y5}, lw_{x6}, lw_{y6}$$
$$rw_{x1}, rw_{y1}, rw_{x2}, rw_{y2}, rw_{x3}, rw_{y3}, rw_{x4}, rw_{y4}, rw_{x5}, rw_{y5}, rw_{x6}, rw_{y6}$$

with:
(m_{xi}, m_{yi}) are the Cartesian coordinates of the 4 landmarks positioned on "the mouth"; (le_{xi}, le_{yi}) are the Cartesian coordinates of the 4 landmarks positioned on "the left eye"; (re_{xi}, re_{yi}) are the Cartesian coordinates of the 4 landmarks positioned on "the right eye"; (lw_{xi}, lw_{yi}) are the Cartesian coordinates of the 6 landmarks positioned on "the left eyebrow"; (rw_{xi}, rw_{yi}) are Cartesian coordinates of the 6 landmarks positioned on "the right eyebrow".

After feature extraction and normalization, in our proposed FER system, we integrated a selection approach to select the best relevant features. There are three main approaches to feature selection commonly discussed in the literature: "Filter" approaches, "Wrapper" approaches and "Embedded" approaches [28]-[30]. Filter-based approaches, which select features independently of any classifier, evaluate features based on their statistical properties or correlation with the output variables without

considering the interaction with the classifier. These approaches provide a quick and easy way to eliminate irrelevant or redundant features before applying more complex modeling techniques. Wrapper-based approaches involve selecting an optimal sub-set of features by directly minimizing the classification error. Unlike filter approaches, Wrappers approaches evaluate feature sub-sets by using a specific classifier, thus accounting for the interaction between feature sub-sets and the classifier's performance. This iterative process allows for more accurate selection of features that directly contribute to the classifier's performance. Embedded approaches integrate feature selection directly into the model training process, simultaneously optimizing both model parameters and feature selection [5]-[31]. In this work, we propose the use of "Wrapper" approach that can select the most relevant features that are more robust to variations such as changes in angle, illumination and occlusion present in real data [32], as well as inter-individual differences in emotional expression [33]-[34]. In fact, feature relevance can help resolve issues like insufficient feature extraction and inaccurate class representation. It can also mitigate the problem of relying on a single similarity measure by allowing a more exhaustive exploration of the feature space and selecting the feature sets best suited to the specific classification task [35].

"Wrapper" approach can be used dynamically to select the best features at each stage of model training, allowing the model to continually adapt to new data and inter-domain variations that may arise, thus increasing its robustness [36]. In the context of a search strategy, this approach involves selecting an algorithm that searches for the optimal sub-set of features based on a given objective function. This transforms the feature-selection problem into an optimization problem, where the goal is generally to maximize classification accuracy while reducing the size of the corresponding feature sub-set. There are many search strategies that can be used in the "Wrapper" selection approach. These can be grouped into three categories: (i) exponential-complexity strategies, (ii) population-based approach strategies and (iii) sequential-selection strategies [37]. In our study, we introduced a selection approach which combines the principle of exponential-complexity strategies with that of sequential selection strategies. Specifically, we evaluate every probable sub-set of features to pinpoint the optimal sub-set. At the same time, we utilize Sequential Forward Selection (SFS), an iterative method that systematically incorporates features from a candidate set. This dual strategy harnesses the robustness of exponential methods to achieve optimal solutions and leverages the efficiency of sequential strategies to enhance the feature-selection process. In what follows, we will give the principle of our proposed selection algorithm. The flowchart depicted in Figure 4 shows the first three iterations of the feature-selection process integrated in the proposed FER system.

As depicted in this figure, the process begins with an empty set devoid of any features. In each iteration, every feature from the initial set is tested to assess its impact on the system's overall performance. Subsequently, the best-performing feature is added to the empty set and the process continues by testing combinations of this resultant set with the remaining features. The general procedure of this process can be given as follows:

Step 1: Initialization
Parameter Set: Start with an empty set of features, '*S*', to be selected from the initial set of features, *F*';
Set $F \leftarrow$ "Initial set of n features";
$S \leftarrow$ "Empty subset";
Basic Models: KNN, DT, NB, LDA;
Performance Criterion: Accuracy maximization;

Step 2: First feature to be selected

1. Test each feature individually:
   - Train the model with each feature individually;
   - Evaluate the model's performance for each case:

   $\forall f_i \in F$, calculate the accuracy using f_i;
   - Select the feature, '$fs1$', that gives the best performance;
   - Add the best feature 'f_{s1}' to the sub-set 'S' and remove it from the set '*F*';
     $F \leftarrow F - \{f_{s1}\}$;
     $S \leftarrow \{f_{s1}\}$.

Step 3: Progressive feature Addition

1. Combination with the best feature:
   - For each feature f_i in set F, combine it with the sub-set S;
   - Train the model on the new $S \cup \{f_i\}$ set;
   - Evaluate performance and add the feature that yields the greatest improvement;
2. Repetition of the feature-selection process until all features are tested:
   - Continue adding features one at a time, testing each combination, until all features have been tested.

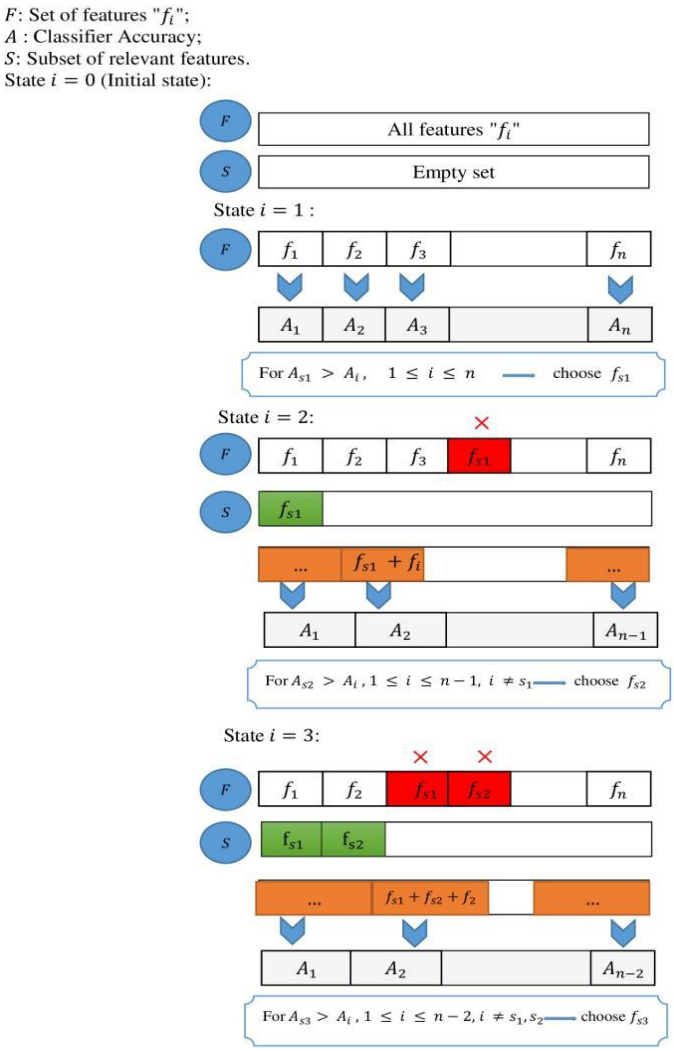


Figure 4. The first three iterations of the feature-selection process integrated in our proposed FER system.

Note here that our proposed Wrapper-based approach carefully selects features during the training phase and applies these same feature columns to the test set without reselecting features for the latter. This approach ensures that the performance metrics obtained during testing reflect the true generalizability of the selected features and the trained model. Consistency in the selection and application of feature columns aims to provide reliable and accurate predictions for unseen data, thereby reinforcing the validity of our method for real-time applications and diverse databases.

For the classification task, we have used KNN, NB, LDA and DT classifiers to find the optimal recognition performance. Although the study in [11] reports that LDA performs better, we decided to test various classifiers, because the effectiveness of our feature-selection approach is directly related to the classifier's performance. Indeed, by systematically eliminating irrelevant or redundant features, our approach could achieve better performance in terms of system response time while maintaining relatively acceptable recognition accuracy when using classifiers other than LDA. Therefore, evaluating the proposed approach using all the classifiers has two advantages:

1. It validates the results reported in [11];
2. It ensures a thorough evaluation of the classifiers to determine the best combination, potentially leading to a more efficient and responsive FER system.

Although most recent work in the field of facial-expression recognition focuses on deep learning-based approaches, our work centers on traditional classification models. The main objective is to assess to what extent our "Wrapper" feature-selection approach can enhance the performance of the recognition system while exploring different model types. Once we have demonstrated the effectiveness of this approach across various classifiers, we can consider generalizing its use to deep learning-based approaches.

In the following section, we will show the contribution of this method to the proposed FER system, by comparing the results obtained before and after its addition.

4. EXPERIMENTAL RESULTS AND DISCUSSION

To test the performance of the proposed modified FER system, it is evaluated on the MUG database. The latter one, proposed by [38], consists of image sequences of 86 subjects (35 women and 51 men), aged from 20 to 35 years and of Caucasian origin, performing facial expressions. Our experimental paradigm is carried out in person-independent mode where no subjects used for testing appear in the training set. Experiments were performed based on 1260 images (180 images for each facial expression), 840 for the training phase (560 images of men and 280 images of women) and 420 for the tests (280 images of men and 140 images of women).

Regarding the work environment, Table 2 outlines the key specifications and software setup of the system used for our work.

Table 2. Technical specifications and environment.

Software/Hardware	**Specification details**
System	Windows 10 Professional, 64-bit
Processor	Intel (R) Core (TM) i5-4300U at 2.49 GHz
RAM	4 GB
Environment	MATLAB R2023a
Toolbox used	MATLAB's Statistics and Machine Learning Toolbox

The analysis and discussion of the obtained results are divided into three sub-sections, presented below.

4.1 The Impact of Adding a Feature-selection Approach on FER System Performance

As explained earlier in the principle of the proposed FER system, after applying the “Wrapper” feature-selection approach to each feature vector (resulting from the extraction of local features for each image of the database using the ASM method), four classification models are used to classify test images into one of seven universal facial-expression classes. The results, obtained using the most relevant features selected by the KNN, DT, NB and LDA classifiers are presented in Tables 3, 4, 5 and 6, respectively. Here, the performance metric used to evaluate the classification performance is the accuracy, which represents the number of correctly predicted data items among all the data. It is given by:

$$Accuracy = \frac{\sum_{i=1\ldots n, j=1\ldots n}[\text{prediction}(i,j)]|i=j}{\sum_{i=1\ldots n, j=1\ldots n}[\text{prediction}(i,j)]} \times 100 \quad (2)$$

with: i: the predicted class; j: the correct class; n: the number of classes.

The classification, using the KNN classifier, was carried out by measuring the Euclidean distance with different values of the nearest neighbor K.

As shown in Table 3, for the KNN classifier, the best accuracy, with a percentage equal to 85.48%, is obtained for K = 1. The classification results, obtained using the DT classifier (according to different minimum numbers of leaf-node observations), reached a maximum accuracy of 70.48% for a minimum leaf size equal to 3 (see Table 4). Similarly, an accuracy of 69.29% is obtained using the NB classifier with Normal (Gaussian) distribution (see Table 5). Finally, the classification results using LDA (for three discriminant types; namely, linear, diagLinear and diagQuadratic) are shown in Table 6. As

shown in this table, the best result, with an accuracy of 100%, is obtained using a linear discriminative function.

Table 3. Classification results using the KNN classifier.

***K* value**	**Maximal Accuracy (%)**	
	Silva et al. FER system [11]	**Our proposed FER system**
1	75.23%	**85. 48%**
3	76.42%	**83. 57%**
5	75.71%	**83. 57%**
7	75.23%	**84. 29%**
9	75.95%	**83. 33%**

Table 4. Classification results using the DT classifier.

Minimum number of branch node observations	**Maximal Accuracy (%)**	
	Silva et al. FER system [11]	**Our proposed FER system**
3	58.09%	**70. 48%**
5	58.80%	**68. 09%**
8	58.33%	**66. 43%**
10	58.57%	**69. 52%**
15	58.10%	**70%**

Table 5. Classification results using the NB classifier.

Distribution	**Maximal Accuracy (%)**	
	Silva et al. FER system [11]	**Our proposed FER system**
Normal	58.33%	**69. 29%** (with 16 features)
Kernel	58.57%	**69. 28%** (with 18 features)

Table 6. Classification results using the LDA classifier.

Discriminant Type	**Maximal Accuracy (%)**	
	Silva et al. FER system [11]	**Our proposed FER system**
Linear	99.71%	**100%**
diagLinear	62.34%	**70. 48%**
diagQuadratic	60.07%	**69. 29%**

A summary of the comparison of the best results between the FER system of Silva et al. [11] and the FER system proposed in this work for the four abovementioned classifiers is shown in Table 7. As shown in this table, the proposed modified FER system provides better performances in all cases. In fact, as illustrated in the same table, by integrating the "Wrapper" feature-selection approach into the traditional system, the proposed modified system achieves an improvement in accuracy between 0.3% and 12%. Moreover, it reduces in the length of the selected feature sub-set which results in a reduction in the number of features between 50% and 77.08%, compared to the original system, for almost identical accuracy or even better.

Simulations utilizing MATLAB tool are performed to evaluate the response time corresponding to the four FER classifiers-based systems before and after incorporating our proposed feature-selection approach. As shown in Table 8, knowing that the NB classifier has the greatest response time (which represents 100%), the results demonstrate the considerable improvement in response time for all evaluated classifiers. Specifically, the KNN classifier's response time decreased from 58.82% to 4.31%, followed by the DT with response time of 4.73% after integrating our feature-selection approach. Similar response times are approximately observed for NB (5.66%) and LDA (4.63%), respectively, illustrating that reducing data complexity through feature selection can significantly speed up classification processes while maintaining comparable performance levels.

Table 7. Performance comparison between the FER system of Silva et al. [11] and the proposed one.

Classifier	Silva et al. FER system [11]		Our proposed FER system		Improvement in accuracy	Rate of reduction in the number of features
	Accuracy (%)	Feature Dimension	Accuracy (%)	Feature Dimension		
KNN	76.42%	48	**85. 48%**	**16**	**9%**	**66. 67%**
DT	58.57%		**70. 48%**	**11**	**11. 91%**	**77. 08%**
NB	58.33%		**69. 29%**	**16**	**10. 96 %**	**66. 67%**
LDA	99.71%		**100%**	**24**	**0. 29%**	**50%**

Table 8. Time-response comparison between the FER system of Silva et al. [11] and the proposed one.

Classifier	Response Time in % (NB: 100%)	
	Silva et al. FER system [11]	Our proposed FER system
KNN	58.82%	**4. 31%**
DT	67.49%	**4. 73%**
NB	100%	**5. 66%**
LDA	74.15%	**4. 63%**

Figure (5) reveals how our proposed FER system outperforms, in terms of accuracy and number of processed features, the original system [11]. In fact, the increase in accuracy can be clearly observed with a reduction in the number of features used in the recognition process. Therefore, we can conclude that feature selection helps improve the recognition accuracy through the removal of redundant and irrelevant features, while also reducing the classifier training time.

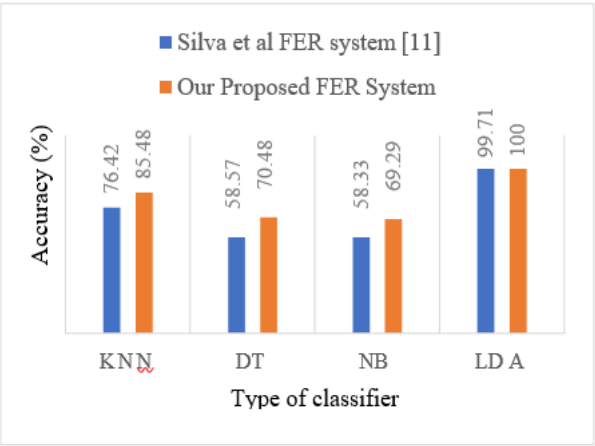


Figure 5. (a) Comparison of the accuracy between the FER system of Silva et al. [11] and the proposed system.

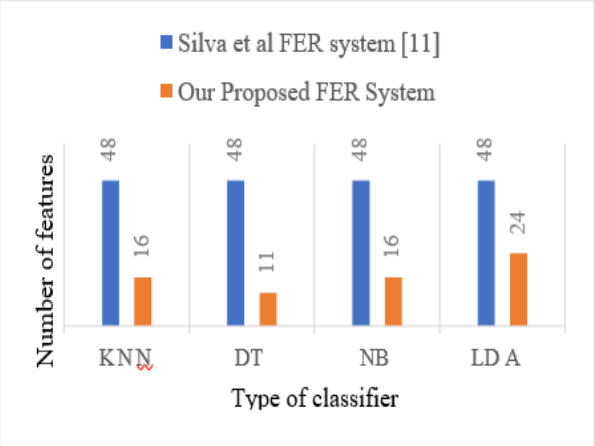


Figure 5. (b) Comparison of feature dimensions between the FER system of Silva et al. [11] and the proposed system.

4.2 Analysis of the Curse of Dimensionality Phenomenon

The term "curse of dimensionality" was introduced, for the first time, by Richard E. Bellman [39] to describe the phenomenon of exponentially increasing the amount of data with dimensionality. From a theoretical point of view, increasing the data dimension adds more information and therefore improves performance. However, in practice, this principle is not always valid. Indeed, in most cases, increasing the dimension leads to increased noise and redundancy during the analysis operation.

In what follows, we will study the impact of this phenomenon on the results of the original FER system. In addition, we will see how the "Wrapper" selection approach can eliminate this phenomenon and thus improve its overall performance. The resulting variations in accuracy, as a function of the number of features, for the four classifiers are shown in Figure (6). In this figure, we represent the response of the FER system, based on each selected feature sub-set Sj (j is the iteration number varying from 1 to 48), using each of the four classifiers.

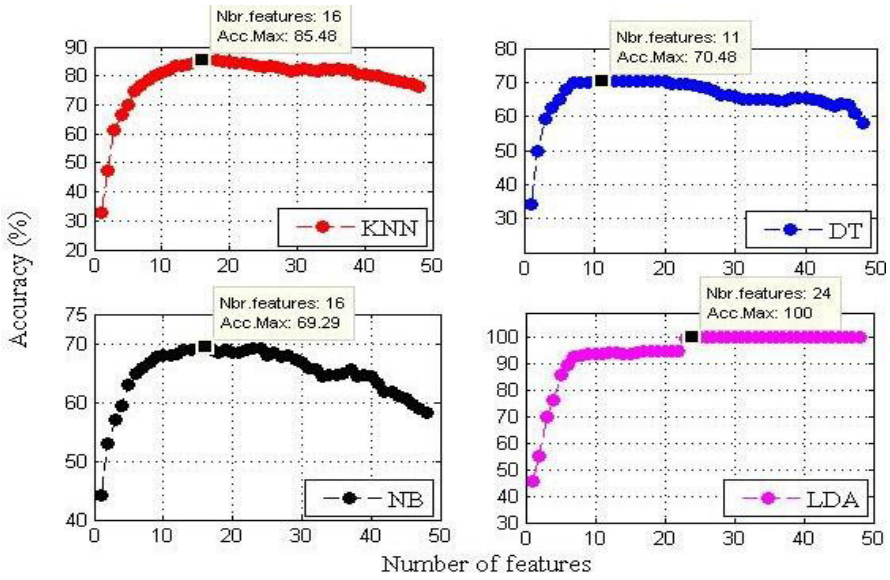


Figure 6. Variation in the accuracy as a function of the number of features:
Top left: KNN classifier ($K = 1$), **Top right:** DT classifier (Number of branch node observations equal to 3), **Bottom left:** NB classifier (Distribution: Normal), **Bottom right:** LDA classifier (Discriminant type: Linear).

As shown in this figure, for the KNN, DT and NB classifiers, each of the corresponding accuracy curves reaches its maximum value, for a specific number of features and then decreases again. This behavior is known as the curse of dimensionality phenomenon or peaking phenomenon. The approach of selecting relevant features makes it possible to overcome this problem and, consequently, improve the performance of the FER system. In fact, as shown in this figure, the FER system, based on the KNN classifier, achieves its maximum accuracy (equal to 85.48%) by exploiting only 16 features. For the DT classifier, this same system achieved a maximum accuracy of 70.48% using only 11 features. Finally, for the NB and LDA classifiers, the FER system achieved values of 69.29% and 100%, respectively, using 16 and 24 features, respectively. For the LDA classifier, as illustrated in this figure, the curve reaches its plateau at 24 selected features with an accuracy of 100%. Therefore, this result demonstrates how the FER system exhibits better performance following the use of the "Wrapper" selection approach and the LDA classifier.

Table 9 provides more details on the feature-selection operation. In particular, it presents the first ten relevant features as well as the recognition accuracy values, corresponding to the different sub-sets Sj, obtained using the LDA classifier. As shown in this table, the selection approach achieved an accuracy of approximately 93% with a dimensional reduction in the feature vector of 83% (8/48), which proves its effectiveness for facial expression-recognition process.

Table 9. The first ten selected features "fsi" and their accuracy values obtained using the LDA classifier for the optimal case.

$Accuracy\ max = 100\%$ obtained using an optimal number of selected features equal to 24.										
i	1	2	3	4	5	6	7	8	9	10
f_{si}	rw_{y5}	m_{y3}	re_{x4}	rw_{y6}	rw_{x2}	rwx_{5}	m_{y1}	m_{y2}	m_{y4}	le_{x4}
Sub-set (Si)	rw_{y5}	S_1+m_{y3}	S_2+re_{x4}	S_3+rw_{y6}	S_4+rw_{x2}	S_5+rw_{x5}	S_6+m_{y1}	S_7+m_{y2}	S_8+m_{y4}	S_9+le_{x4}
Accuracy (%)	45.48%	55.24%	69.52%	75.95%	85.71%	89.05%	92.38%	93.10%	93.57%	93.57%

4.3 Relevance of Local Facial Regions on FER-system Performance

To determine the most important local region of a face, the most relevant local features, allowing the system to achieve maximum accuracy, must be studied.

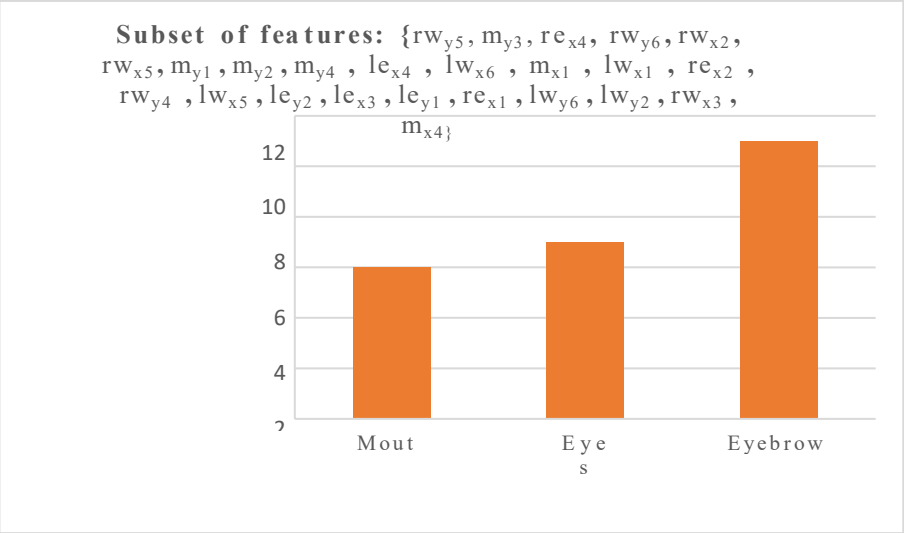


Figure 7. Most relevant features according to their position on the face using LDA classifier.

As shown in Figure (7), after an analysis of the most relevant features (according to their positions on the face), we find that most of them belong to the local "eyebrow" region (with 11 features/24), followed by the "eyes" region (with 7 features/24) and finally the "mouth" region (with 6 features/24). This result shows the importance of the "eyebrows" region in the field of facial-expression recognition. In fact, eyebrows are highly mobile and their movement significantly contributes to expressing various emotions. The position and movement of eyebrows provide subtle cues about a person's intentions and internal state, often preceding changes in other aspects of the face. This nuanced expressiveness underscores the crucial role of eyebrows in non-verbal communication, making them an essential element for understanding and accurately recognizing facial expressions. Despite their importance, this concept has been overlooked by researchers in the field. Typically, studies focus primarily on the three main facial regions: the mouth, nose and eyes. However, future research could enhance algorithms by incorporating eyebrow dynamics, potentially leading to more accurate facial-expression recognition.

4.4 Comparison with State-of-the-Art Methods

When we compare our approach to those proposed in [15], [40]-[44], we observe that each feature-selection approach has its strengths and weaknesses. The choice of a specific approach often depends on the specific application context and practical constraints, such as the trade-off between precision and acceptable computational complexity.

- Our approach demonstrates optimal performance in terms of accuracy by capturing complex interactions between features. This is particularly effective when the number of features is not too high as the computational cost remains moderate. Consequently, our method is particularly suitable when computational resources are available and achieving the highest precision is crucial;
- Relief-F and Minimum Redundancy Maximum Relevance (mRMR) are well suited for managing feature redundancy and noise while being less computationally expensive. They strike a balance between relevance and redundancy, making them efficient for large databases with many features;
- Chi-square, GR and IG are fast and straightforward to implement. However, they may lack the sophistication needed to capture complex interactions between features, which could lead to sub-optimal performance in scenarios where such interactions are critical;
- The SCA and Autoencoder methods offer global exploration capabilities and the ability to capture nonlinear relationships, respectively. Despite their advanced capabilities, these methods require fine-tuning and are computationally intensive, which might limit their practicality for large- scale problems without sufficient computational resources.

Table 10 compares the performance of our proposed FER system to those of the most relevant systems reported in the literature.

Table 10. Comparison between the performance of the proposed FER system and those of the most relevant reported in the literature.

Ref.	Database	Region of interest	No. of emotions	FS Method	Type of approach	Classifier	Accuracy in %
Our work	MUG	Mouth, Eyes and Eyebrows	7	Our proposed FS approach	Wrapper	LDA	**100%**
[3]	CK+	Whole face	8	Relief-F	Filter	KNN	94.93%
[4]	CK+	Whole face	8	Relief-F	Filter	KNN	94.93%
[5]	SFE	Whole face	8	SCA	Wrapper	KNN	97.80%
[6]	TFEID	Whole face divided into regular regions	8	Pairwise FS	Hybrid	SVM	99.63%
[7]	CK+	Left eye regions and half mouth region	8	GR	Filter	KNN	91.01%
[8]	JAFFE	Whole face	7	RV Correlation	Filter	SVM	96.53%
[9]	JAFFE	Whole face	7	IUTWSVM	Embedded	SVM	91.43%
[10]	Xceptio	Whole face	6	Autoencoder	Embedded	DCNN	95.23%

The analysis realized herein reveals that the proposed system, which selects regions of interest (ROIs) such as the mouth, eyes and eyebrows, achieves a remarkable accuracy of 100% using the MUG database. Unlike whole-face analysis, focusing on these key regions makes it possible to better capture emotional nuances, thus improving the robustness and efficiency of the models. Finally, this comparison emphasizes the promising effectiveness of the proposed method, even amidst the variety of existing approaches in this field.

5. CONCLUSION

In this paper, an improved FER system, based on the integration of "Wrapper" selection approach and the use of relevant information provided by local regions of the face, is proposed. This FER system, which is realized in several steps, can be used to achieve more efficient recognition of several facial expressions, namely, happiness, anger, sadness, surprise, disgust, fear and neutral. In its first step, which consists of feature extraction and normalization, ASM was used to extract facial features and GPA was used to normalize the extracted features. Concerning its second step, intended for the selection of features, a "Wrapper" approach was integrated to choose, among the features already extracted, only those that were the most relevant. Here, to determine the recognition accuracy, several classifiers, specifically KNN, DT, NB and LDA, were used. Simulation results, based on the MUG database demonstrated that the proposed FER system outperforms the traditional system in terms of accuracy and response time. Indeed, our approach effectively addresses the curse of dimensionality phenomenon for all four classifiers. Furthermore, the results highlighted that the majority of relevant extracted local features belong to the local "eyebrow" region, underscoring its importance in decoding emotions through facial expressions. Eyebrows play a crucial role in expressive movements, intentional cues and the accentuating facial features, which may explain their significance in FER systems. Recognizing their increasing importance in future research could lead to advancements in recognition algorithms, human- machine interactions and applications in various fields, such as mental health, security and marketing. Integrating this dynamic in FER systems and emotional analysis could provide a richer and more precise understanding of human-machine interactions.

Our approach effectively mitigates challenges in recent FER systems, such as overfitting and handling expression-unrelated variations like lighting and head pose, thereby enhancing accuracy and robustness. However, relying on features extracted by the Cootes and Taylor algorithm may limit the system's ability to capture the full spectrum of facial expressions and nuances. Additionally, the MUG database's constraints, including small sample size, controlled capture environment and lack in ethnicity and age, may limit the generalizability of our findings.

In future work, we aim to integrate deep-learning techniques for feature extraction to enhance captured expressions. We also plan to expand and diversify our database to improve the system's generalization to more varied and naturalistic settings.

REFERENCES

[1] S. Li and W. Deng, "Deep Facial Expression Recognition: A Survey," IEEE Transactions on Affective Computing, vol. 13, no 3, pp. 1195-1215, 2020.

[2] J. F. Cohn and F. De la Torre, "Automated Face Analysis for Affective Computing," The Oxford Handbook of Affective Computing, pp. 131-150, 2015.

[3] M. B. Abdulrazaq, M. R. Mahmood, S. R. Zeebaree, M. H. Abdulwahab, R. R. Zebari and A. B. Sallow, "An Analytical Appraisal for Supervised Classifiers' Performance on Facial Expression Recognition Based on Relief-F Feature Selection," Journal of Physics: Conference Series, IOP Publishing, vol. 1804, no. 1, p. 012055, February 2021.

[4] M. R. Mahmood, "Two Feature Selection Methods Comparison Chi-Square and Relief-F for Facial Expression Recognition," Journal of Physics: Conference Series, IOP Publishing, vol. 1804, no. 1, pp. 012056, February 2021.

[5] A.A. Ewees, H. A. ElLaban and R. M. ElEraky, "Features Selection for Facial Expression Recognition," Proc. of the 2019 10th IEEE Int. Conf. on Computing, Communication and Networking Technologies (ICCCNT), pp. 1-6, Kanpur, India, July 2019.

[6] M. J. Cossetin, J. C. Nievola and A. L Koerich, "Facial Expression Recognition Using a Pairwise Feature Selection and Classification Approach," Proc. of the 2016 IEEE Int. Joint Conf. on Neural Networks (IJCNN), pp. 5149-5155, Vancouver, BC, Canada, July 2016.

[7] H. I. Dino and M. B. Abdulrazzaq, "A Comparison of Four Classification Algorithms for Facial Expression Recognition," Polytechnic Journal, vol. 10, no. 1, pp. 74-80, 2020.

[8] Y. Wang, Y. Lu and X. Wan, "Multi-feature Fusion Based on RV Correlation Coefficient for Facial Expression Recognition," Journal of Physics: Conference Series, IOP Publishing, vol. 2562, no. 1, p. 012027, August 2023.

[9] B. Richhariya and D. Gupta, "Facial Expression Recognition Using Iterative Universum Twin Support Vector Machine," Applied Soft Computing, vol.76, pp. 53-67, 2019.

[10] F. M. Talaat, Z. H. Ali, R. R. Mostafa and N. El-Rashidy, "Real-time Facial Emotion Recognition Model Based on Kernel Autoencoder and Convolutional Neural Network for Autism Children," Soft Computing, vol. 28, pp. 6695-6708, 2024.

[11] C. Silva, A. Sobral and R. T. Vieira, "An Automatic Facial Expression Recognition System Evaluated by Different Classifiers," Proc. of the X Workshop de Visão Computacional, Uberlândia, Minas Gerais, Brazil, pp. 208-212, October 2014.

[12] T. F. Cootes, G. J. Edwards and C. J. Taylor, "Comparing Active Shape Models with Active Appearance Models," Proc. of the British Machine Vision Conference (BMVC), vol. 99, no. 1, pp. 173-182, September 1999.

[13] J. Li, K. Cheng, S. Wang, F. Morstatter, R. P. Trevino, J. Tang and H. Liu, "Feature Selection: A Data Perspective," ACM Computing Surveys (CSUR), vol. 50, no. 6, pp. 1-45, 2017.

[14] G. Chandrashekar and F. Sahin, "A Survey on Feature Selection Methods," Computers & Electrical Engineering, vol. 40, no. 1, pp. 16-28, 2014.

[15] J. R. Quinlan, "Induction of Decision Trees," Machine Learning, vol. 1, pp. 81-106, 1986.

[16] Y. Saeys, I. Inza and P. Larranaga, "A Review of Feature Selection Techniques in Bioinformatics," Bioinformatics, vol. 23, no. 19, pp. 2507-2517, 2007.

[17] P. Viola and M. J. Jones, "Robust Real-time Face Detection," International Journal of Computer Vision, vol. 57, no. 2, pp. 137-154, 2004.

[18] X. Zhao and S. Zhang, "A Review on Facial Expression Recognition: Feature Extraction and Classification," IETE Technical Review, vol. 33, no. 5, pp. 505-517, 2016.

[19] P. Ekman, "Facial Expressions," Handbook of Cognition and Emotion, vol. 16, no. 301, e320, 1999.

[20] J. N. Bassili, "Facial Motion in the Perception of Faces and of Emotional Expression," Journal of Experimental Psychology: Human Perception and Performance, vol. 4, no. 3, pp. 373, 1978.

[21] H. Xiong, D. Zhang, C. J. Martyniuk, V. L. Trudeau and X. Xia, "Using Generalized Procrustes Analysis (GPA) for Normalization of cDNA Microarray Data," BMC Bioinformatics, vol. 9, no. 1, pp. 1-13, 2008.

[22] T. Cover and P. Hart, "Nearest Neighbor Pattern Classification," IEEE Transactions on Information Theory, vol. 13, no. 1, pp. 21-27, 1967.

[23] D. D. Lewis, "Naive (Bayes) at Forty: The Independence Assumption in Information Retrieval," Proc.

of the European Conf. on Machine Learning, pp. 4-15, Berlin, Heidelberg: Springer Berlin Heidelberg, April 1998.

[24] T. Hastie, R. Tibshirani, J. H. Friedman and J. H. Friedman, The Elements of Statistical Learning: Data Mining, Inference and Prediction, vol. 2, pp. 1-758, eBook ISBN: 978-0-387-84858-7, New York: Springer, 2009.

[25] N. Suguna and K. Thanushkodi, "An Improved K-Nearest Neighbor Classification Using Genetic Algorithm," International Journal of Computer Science Issues, vol. 7, no. 2, pp. 18-21, 2010.

[26] P. Xanthopoulos, P. M. Pardalos and T. B. Trafalis, "Linear Discriminant Analysis," Robust Data Mining, Springer, pp. 27-33, Part of the Book Series: Springer Briefs in Optimization, New York, NY, 2013.

[27] B. Charbuty and A. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," Journal of Applied Science and Technology Trends, vol. 2, no. 01, pp. 20-28, 2021.

[28] S. Khalid, T. Khalil and S. Nasreen, "A Survey of Feature Selection and Feature Extraction Techniques in Machine Learning," Proc. of the 2014 IEEE Science and Information Conf., pp. 372-378, London, UK, August 2014.

[29] J. Miao and L. Niu, "A Survey on Feature Selection," Procedia Computer Science, vol. 91, pp. 919-926. 2016.

[30] S. Liu and M. Motani, "Improving Mutual Information Based Feature Selection by Boosting Unique Relevance," arXiv preprint, arXiv: 2212.06143, 2022.

[31] D. Choudhary and J.Shukla, "Feature Extraction and Feature Selection for Emotion Recognition Using Facial Expression," Proc. of the 2020 IEEE 6th Int. Conf. on Multimedia Big Data (BigMM), pp. 125-133, New Delhi, India, September 2020.

[32] Z. Mortezaie and H. Hassanpour, "A Survey on Age-invariant Face Recognition Methods", Jordanian Journal of Computers and Information Technology (JJCIT), vol. 05, no. 02, pp. 87 - 96, August 2019.

[33] M. Jiang and S. Yin, "Facial Expression Recognition Based on Convolutional Block Attention Module and Multi-feature Fusion," International Journal of Computational Vision and Robotics, vol. 13, no. 1, pp. 21-37, 2023.

[34] Y. Jiang and S. Yin, "Heterogenous-View Occluded Expression Data Recognition Based on Cycle-consistent Adversarial Network and K-SVD Dictionary Learning under Intelligent Cooperative Robot Environment," Computer Science and Information Systems, vol. 20, no. 4, pp. 1869-1883, 2023.

[35] X. Meng, X. Wang, S. Yin and H. Li, "Few-shot Image Classification Algorithm Based on Attention Mechanism and Weight Fusion," Journal of Engineering and Applied Science, vol. 70, no. 1, p. 14, 2023.

[36] L. Teng, "Brief Review of Medical Image Segmentation Based on Deep Learning," IJLAI Transactions on Science and Engineering, vol. 1, no. 02, pp. 01-08, 2023.

[37] N. El Aboudi and L. Benhlima, "Review on Wrapper Feature Selection Approaches," Proc. of the 2016 IEEE Int. Conf. on Engineering & MIS (ICEMIS), pp. 1-5, Agadir, Morocco, September 2016.

[38] N. Aifanti, C. Papachristou and A. Delopoulos, "The MUG Facial Expression Database," Proc. of the 11th IEEE Int. Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 10), pp. 1-4, Desenzano del Garda, Italy, April 2010.

[39] R. J. N. J. Bellman, Dynamic Programming, ISBN 0-691-07951-X, Princeton University Press, Princeton, New Jersey, 1957.

[40] I. Kononenko, E. Šimec and M. Robnik-Šikonja, "Overcoming the Myopia of Inductive Learning Algorithms with RELIEFF," Applied Intelligence, vol. 7, pp. 39-55, 1997.

[41] H. Peng, F. Long and C. Ding, "Feature Selection Based on Mutual Information: Criteria of Max-dependency, Max-relevance and Min-redundancy," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 8, p. 1226-1238, 2005.

[42] H. Liu and R. Setiono, "Chi2: Feature Selection and Discretization of Numeric Attributes," Proc. of the 7th IEEE Int. Conf. on Tools with Artificial Intelligence, pp. 388-391, Herndon, VA, USA, 1995.

[43] S. Mirjalili, "SCA: A Sine Cosine Algorithm for Solving Optimization Problems," Knowledge-based Systems, vol. 96, pp.120-133, 2016.

[44] G. E. Hinton and R. R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks," Science, vol. 313, no. 5786, pp. 504-507, 2006.

ملخص البحث:

توفّر أنظمة تمييز تعابير الوجْه طريقةً مهمّةً للتعبير عن الحالة النّفسية والعاطفية للنّاس وتفسيرها. وتعمل هذه الأنظمة على تحويل صورةِ الوجهِ إلى مجموعةٍ من السّمات أو الخصائص لتدريب مُصَنِّفٍ قادرٍ على التّمييز بين أصنافٍ مختلفةٍ من المشاعِر. ومع ذلك، فإنّ المشكلة الّتي تعْترض تلك الأنظمة تكمن في أنّ مُتّجهات السِّمات الّتي يتمّ الحصول عليها قد تحتوي أحياناً على سماتٍ مكرّرة أو ليست ذات علاقة، الأمر الذي يؤدي إلى تقليل دقّة المصنِّف بالإضافة إلى زيادة زمن الحساب. وللتّغلُّب على هذه المشكلة، يجب تقليل عدد السِّمات من خلال انتقاء السِّمات ذات العلاقة الأقوى دون غيرها.

في هذه الورقة، نعمل على دراسةِ أثر إضافة انتقاء السِّمات بطريقة "اللّفّ" واستخدام المعلومات التي توفرها مناطق محلية من الوجْه، مثل: الفم، والعينين، والرّموش، ودراسة أثرها على أداء الأنظمة التّقليدية لتمييز تعابير الوجه. والهدف هنا هو فحص وتحليل كيف يمكن لهذه التّوليفة أن تحسّن من أداءِ النّظام الأصلي التّقليدي.

وتبيّن النّتائج أنّ نظام تمييز تعابير الوجه المعزَّز باستراتيجية انتقاء السِّمات المقترحة أعطى نتائج أفضل مقارنةً بالنّظام الأصلي لجميع المصنِّفات المستخدمة. وفي الواقع، لوحظ الحصول على تقليلٍ معتبَر لعدد السِّمات (حتى 50%) ودقّة بلغت (100%) عند استخدام مصنِّف التّحليل التّمايُزي الخطّي (LDA)، الأمر الّذي يُعدّ تحسيناً ملحوظاً من حيث زمن الحسابات والفاعلية و حجم الذّاكرة.

بالإضافة إلى ذلك، كانت غالبية السِّمات المستخدمة جزءاً من "منطقة الرُّموش"، الأمر الّذي يثبت أهمية استخدام معلوماتٍ من مناطق محلية معينة من الوجه في المهامّ المتعلّقة بتمييز المشاعِر.

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 10, No. 04, December 2024.

LOW-PROFILE BI-SLOT DUAL-BAND (2.45/5.76GHZ) ANTENNA FOR WIRELESS COMMUNICATION SYSTEMS

Ahmad Abbas AL Rimi[1*], Asmaa Zugari[1], Mohssine El Ouahabi[2] and Mohsine Khalladi[1]

(Received: 7-May-2024, Revised: 4-Jul.-2024, Accepted: 27-Jul.-2024)

ABSTRACT

This article presents a compact dual-band patch antenna designed for wireless communication systems. The proposed antenna design operates efficiently at both 2.45 GHz and 5.76 GHz achieving high performance despite its compact size. The Defected Microstrip Structure (DMS) is used to create the desired frequency bands and improve the impedance matching at the respective required frequencies, respectively. The antenna is manufactured on a low-cost FR-4 substrate, with dimensions of 27 mm × 27 mm × 1.6 mm. The simulated radiation efficiencies at 2.45 GHz and5.76 GHz are 95% and 73%, respectively. In addition, the gain of the suggested antenna is more than 2.65 dB and around 2.84 dB at the desired frequency bands. The suggested antenna is simulated, manufactured, tested and verified practically. The return losses measured in the lower and upper resonance bands are 29.94% (covering the frequency range from 2.13 to 2.88 GHz) and 6.2% (covering the frequency range from 5.65 to 6.01 GHz), respectively. The radiation pattern is measured and compared with the simulation one. The design is simple, easy to carry on and is very compact, while enabling seamless operation across two different frequency bands, which makes it suitable for diverse wireless applications.

KEYWORDS

Compact, Dual-band, DMS, WLAN.

1. INTRODUCTION

Wireless communication has undergone rapid advances in contemporary society, encompassing a variety of technical fields. Following the FCC's announcement allowing individuals to use medical, industrial and scientific frequencies without the need for a licence, the scientific community has been provided with an important opportunity to create wireless devices suitable for short-range communication. The most notable examples are Bluetooth, which operates in the globally allocated 2.4 GHz ISM band [1] and wireless local area networks (WLANs), which operate at frequencies of 5.2 GHz and 5.8 GHz. WLANs were conceived as adaptable data-communication systems, serving as substitutes for or extensions to wired local networks. Using radio-frequency technology, these networks send and receive data wirelessly in the atmosphere, significantly reducing reliance on wired connections and seamlessly integrating uninterrupted connectivity with user mobility. In today's landscape, wireless LANs are rapidly gaining popularity in diverse sectors, such as healthcare, manufacturing and academia. These industries have reaped substantial benefits from portable devices that enable real-time transfer of data to central processing centres. In addition, wireless local area networks are increasingly recognized as a reliable and cost-effective means of achieving fast wireless internet access. They are being adopted as versatile connectivity solutions in a wide range of applications [2].

The IEEE 802.11 WLAN standards encompass three operational frequency bands: 2.4 GHz (2.400 GHz - 2.484 GHz), 5.2 GHz (5.150 GHz - 5.350 GHz) and 5.8 GHz (5.725 GHz - 5.825 GHz). In the IEEE 802.11a standard, WLANs operate within the higher-frequency range, covering the range of 5.15 GHz to 5.35 GHz and 5.725 GHz to 5.825 GHz. On the other hand, WLANs that adhere to the IEEE 802.11b/g standards use the 2.4 GHz band, which extends from 2.4 GHz to 2.484 GHz [2][3][4]. In the contemporary landscape, dual-band WLAN systems, which include IEEE 802.11a /b /g standards are becoming increasingly popular [5]. Consequently, to meet the demands of wireless communication, it is necessary to develop compact, high-performance antennas capable of operating in the dual bands of 2.4 and 5.8 GHz, while at the same time offering exceptional radiation properties

1. A. AL Rimi (Corresponding Author), A. Zugari and M. Khalladi are with Information Systems and Telecommunications Laboratory, University of Abdelmalek Essaadi, Faculty of Science Tetuan, Morocco. Emails: alrimiahmed75@gmail.com, asmaa.zugari@gmail.com and m_khall@hotmail.com
2. M. El Ouahabi is with National School of Applied Sciences, University of Abdelmalek Essaadi, Email: moelouahabi@uae.ac.ma

[2], [6]. Nonetheless, the design of compact dual-band antennas is a major challenge, given the need to meet wide-bandwidth requirements, as well as to achieve a balanced radiation pattern, constant gain, compact size and simple manufacturing processes [7][8][9]. On the other hand, the narrow-band antenna can be transformed into a multiband antenna by introducing a slit or slot at strategically chosen positions on its patch or ground plane. The dimensions, geometry and position of this slot are crucial to the creation of resonances and band characteristics [10]. A number of dual-band microstrip antennas have been reported in the literature, but they typically exhibit rather large sizes and relatively complex designs. For instance, a dual-band printed slot antenna using the Cantor fractal has achieved two operating frequency bands of 2.35–3.61 GHz and 5.15–6.25 GHz; achieving this performance necessitated large physical dimensions of 50 x 50 mm² [11]. Similarly, an annular Koch snowflake fractal antenna designed for dual-band performance covered 2.24–2.93 GHz and 4.48–5.54 GHz, with dimensions of 40 × 40 mm² [12]. A dual Wide Band Monopole Antenna for WiFi and WiMAX systems, has been designed to cover 2.3–2.98 GHz and 5.13–7.75 GHz, with dimensions of 40 × 44 mm² [13]. A dual Wideband Monopole antenna for GSM/UMTS/LTE/WiFi/and Lower UWB applications consists of two sickle-shaped radiators and a slotted ground plane, achieving operating bands of 1.5–2.8 GHz and 3.2–6 GHz, but it possesses a large size of 57 × 37.5 mm² and a relatively complex design [14]. A DGS antenna designed for medical applications has been published. It operates at two frequency bands with dimensions of 58 x 40 mm² [15]. A dual-band microstrip antenna using a polarization conversion metasurface structure is proposed in [16], with dimensions of 40 × 49 mm². A dual-band semi-circular patch antenna for WiMAX and WiFi-5/6 applications witch covers 2.39–3.75 GHz and 5.39–7.18 GHz, with dimensions of 30 × 40 mm² was proposed in [17]. In [18], a dual-band antenna for WiMax and Wi-Fi applications is introduced with dimensions of 40 × 40 mm². Reference [19] incorporates a Dual-Band Microstrip Patch Antenna Design Using C-shaped slot to cover the 2.4 and 5 GHz frequency bands, with dimensions of 53 × 49 mm².

This paper proposes a compact dual-band patch antenna that operates efficiently at 2.45 GHz and 5.76 GHz, achieving high efficiencies of over 95% and 73% in both operating bands. The antenna attains gains of 2.65 dB and 2.84 dB, respectively. The Defected Microstrip Structure (DMS) technique is used to achieve the desired frequency bands and enhance impedance matching at both frequencies.

The proposed antenna has been simulated, manufactured, tested and practically verified, with results showing good agreement with simulations. These results demonstrate the antenna's suitability for WLAN, ISM and WiMAX applications. Additionally, it is characterized by its small size, ease of manufacture and smooth operation across two different frequency bands.

The design structure, simulated and measured results, analysis and concluding remarks are presented in Sections 2, 3, 4, 5 and 6.

2. ANTENNA DESIGN EVOLUTION

2.1 Design Process

The antenna design process is structured into four stages. Figure 1 depicts the evolution of the proposed antenna, while Figure 2 presents the simulated reflection-coefficient curves corresponding to each stage.

In Step 1, the design of the initial rectangular patch antenna is calculated using the well-known transmission line theory [20]:

$$LP = \frac{C}{2f_r\sqrt{\varepsilon reff}} - 0.824h - \left(\frac{(\varepsilon reff + 0.3)(\frac{W_p}{h} + 0.264)}{(\varepsilon reff + 0.258)(\frac{W_p}{h} + 0.8)} \right)$$

$$W_p = \frac{C}{2f_r\sqrt{\frac{\varepsilon_r + 1}{2}}}$$

$$\varepsilon_{reff} = \frac{\varepsilon_r + 1}{2} + \frac{\varepsilon_r - 1}{2}\left(1 + 12\frac{h}{W}\right)^{-0.5}$$

where, Lp is the length of the patcn, Wp is the width of the patcn, C is the speed of light (3×10^8 m/s), fr is the resonant frequency, ϵr is the substrate relative permittivity and h is the substrate thickness.

In the second stage, enhancing impedance matching with a reduced ground plane leads to a resonant frequency of 5.59 GHz. The third stage introduces a rectangular slot (L2 x W1) on the top of the patch, resulting in a resonant frequency of 2.78 GHz. Finally, in the fourth stage, adding another rectangular open-ended slot (L2 x W1) below the patch produces two resonant frequencies at 2.9 GHz and 6.27 GHz.

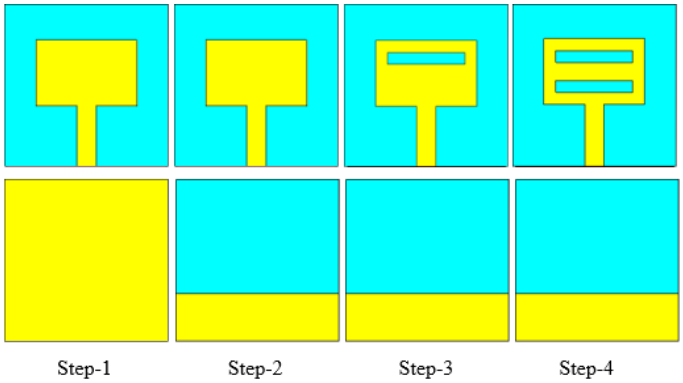


Figure 1. Design procedure of the proposed antenna.

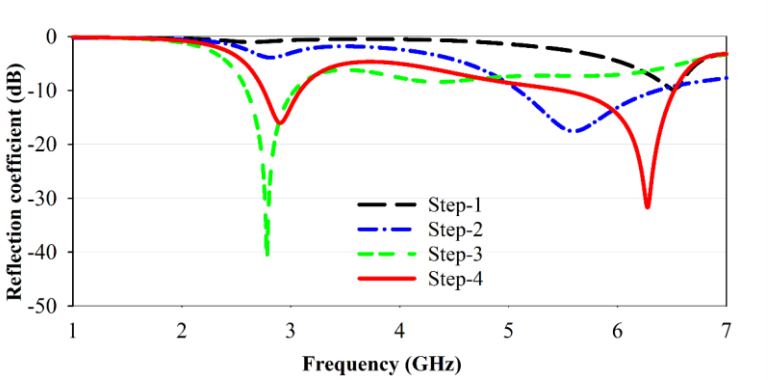


Figure 2. Simulated reflection coefficients for all design steps.

2.2 Final Structure and Dimensions

The design of the proposed antenna and its corresponding prototype are illustrated in Figure 3. The antenna is manufactured on an FR-4 substrate, which has a relative permittivity of 4.3, a height (h) of 1.6 mm and a loss tangent of 0.025. A 50-ohm microstrip line is used to connect the antenna to an external source *via* an SMA connector. Dual-band radiation at the specified frequencies is achieved by drilling two parallel rectangular slots in the patch and utilizing defected ground structure (DMS) technology to improve impedance matching. Both the patch and ground plane are made of copper. The optimal dimensions of the antenna, determined through a parametric study using CST Microwave Studio, are outlined in Table 1. Computer Simulation Technology (CST) is used to design and simulate the proposed antenna.

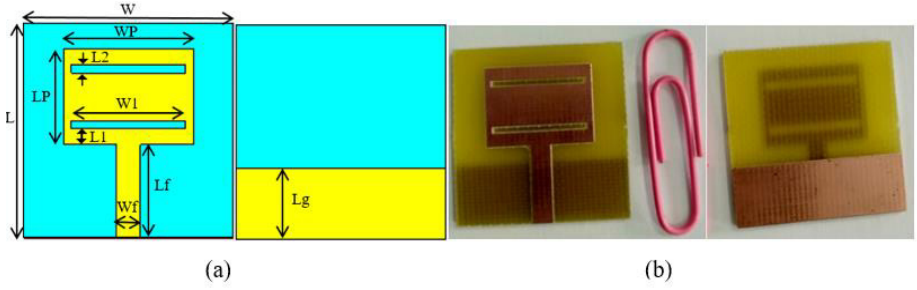


Figure 3. (a) Top and bottom views of the designed antenna geometry, (b) The manufactured prototype.

Table 1. Optimized dimensions of the proposed antenna.

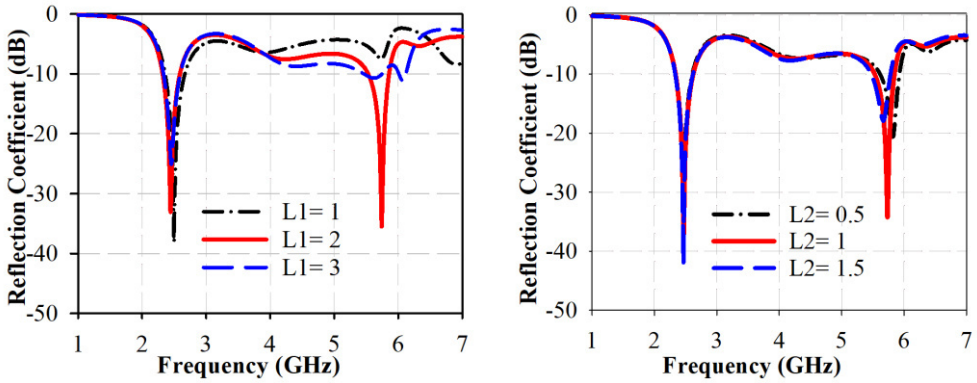
Parameters	L	W	LP	WP	Lf	Wf	Lg	L1	L2	W1	h
Value(mm)	27	27	12	16.7	11.7	3.137	9	2	1	14.7	1.6

3. PARAMETRIC STUDIES

The main purpose of this section is to study the effect of antenna parameters on resonant frequencies and matching. This analysis aims to determine the optimal values for the two rectangular slots etched inside the patch using the DMS technique. This is achieved by adjusting a single parameter while maintaining others constant.

3.1 Effect of L1 and L2

As shown in Figure 4(a), modifications in the position of the two rectangular slots from L1 = 1 mm to L1 = 3 mm leads to a decrease in return loss at the lower frequency of 2.45 GHz, while at both L1 = 1 mm and L1 = 3 mm, the upper frequency of 5.76 GHz is entirely absent. Moreover, as depicted in Figure 4(b), altering the width of the rectangular slot from L2= 0.5 mm to L2= 1.5 mm results in a decrease in return loss at the higher frequency. Additionally, this modification causes a shift in the upper frequency from 5.82 GHz to 5.66 GHz up to 5.76 GHz, while the lower frequency of 2.45 GHz remains unaffected. It is observed that the optimal outcome is achieved for L1= 2 mm and L2= 1 mm in both frequency bands.

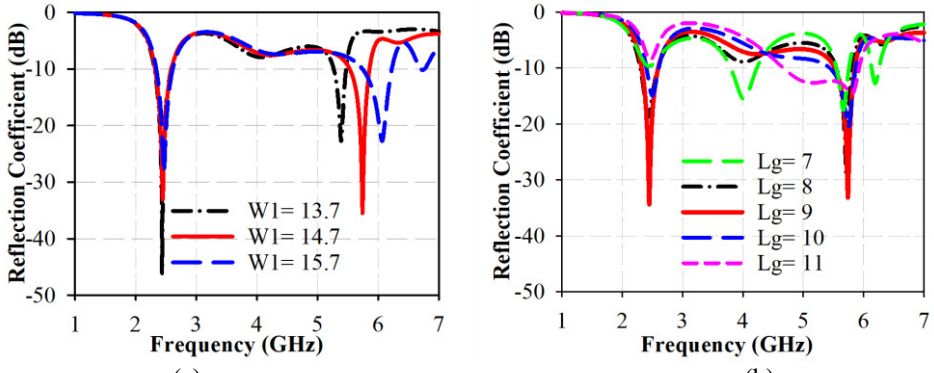


(a) (b)

Figure 4. Simulated S_{11} parameter for various values of (a) L1 and (b) L2.

3.2 Effect of W1 and Lg

Figure 5(a) shows the simulated reflection coefficient of W1; when W1 is increased from 13.7 mm to 15.7 mm, the upper resonant frequency decreases progressively from 6.15 GHz to 5.76 GHz and finally to 5.4 GHz. Our objective is to achieve a resonant frequency of 5.76 GHz, the optimal value is observed at W1 = 14.7 mm. Also, as observed in Figure 5(b), the simulated reflection coefficient decreases when the Lg value increases or decreases from 9 mm, indicating that 9 mm is the optimal value.



(a) (b)

Figure 5. Simulated S_{11} parameter for different values of (a) W1 and (b) Lg.

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 10, No. 04, December 2024.

4. Simulation Results and Discussion

The designed antenna is manufactured using the LPKF ProtoMat E33 milling machine. The Rohde and Schwarz ZVB 20 Vector Network Analyzer is utilized to measure the S-parameters, as depicted in Figure 6(b). It is evident that there is acceptable agreement between the measurements and the simulation results as can be seen in Figure 6(a). The impedance bandwidths (S_{11} < -10 dB) of the suggested antenna are (2.13–2.88 GHz), corresponding to a range of 29.94% (lower-frequency band) and (5.65–6.01 GHz), corresponding to a range of 6.2% (higher-frequency band), respectively. Figure 7(a) shows the VSWR graph at the center frequencies for the proposed antenna, with values less than 2. Specifically, the designed antenna has a VSWR of 1.06 at 2.45 GHz and 1.11 at 5.76 GHz, indicating favorable impedance matching. Figure 7(b) presents the Z-parameter of the antenna, demonstrating that the impedance is approximately 50 Ohms at both 2.45 GHz and 5.76 GHz, confirming that the antenna provides perfect impedance matching.

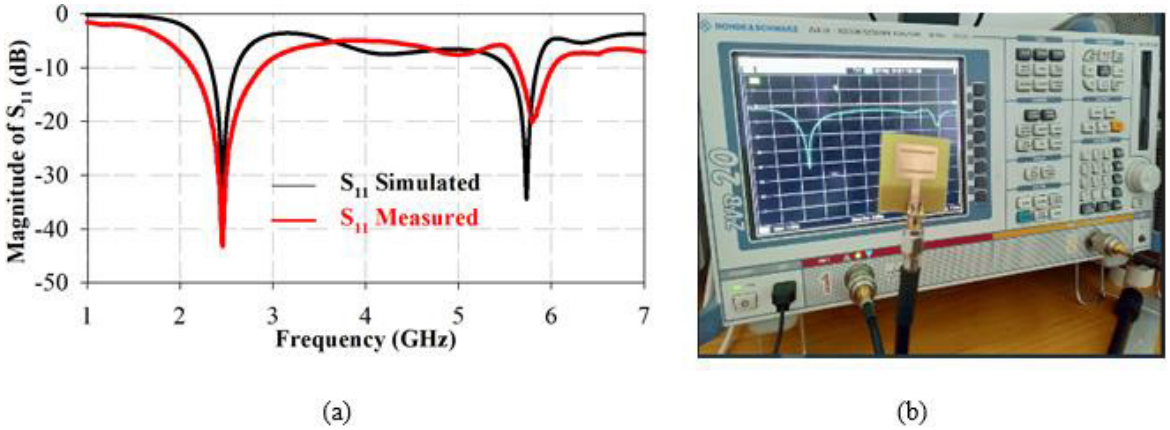


Figure 6. (a) Measured and simulated magnitude of S_{11}, (b) Measurement setup.

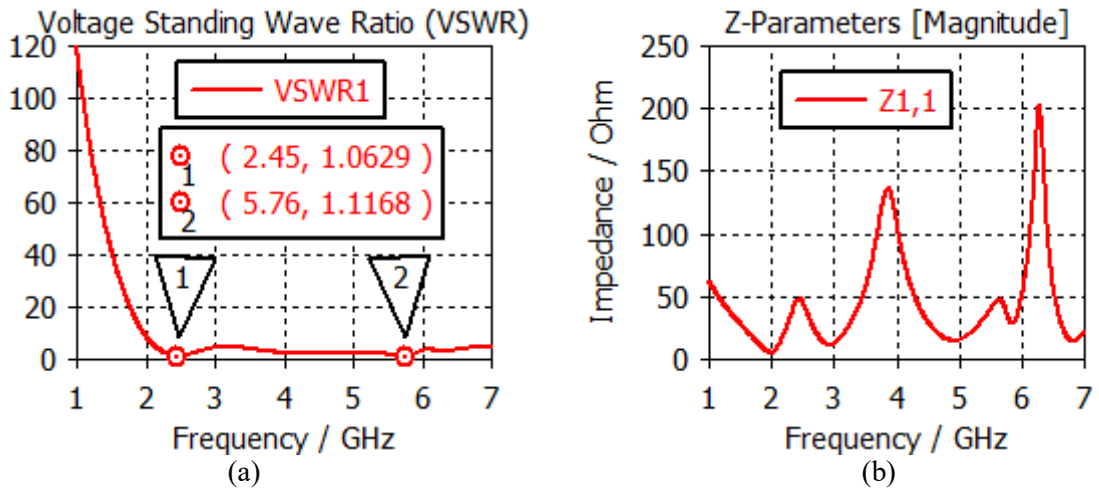


Figure 7. (a) VSWR of proposed antenna and (b) Z- parameters.

4.1 Surface-current Distribution Analysis

To evaluate the performance of the proposed antenna, we conducted simulations of the surface-current distribution at two resonant frequencies, 2.45 GHz and 5.76 GHz (Figure 8). The results exhibit distinct patterns at each frequency, revealing specific operational characteristics of the antenna.

At 2.45 GHz: The surface current predominantly gathers around the feed point, slots and antenna edges. This concentration signifies efficient coupling and radiation at the fundamental resonant frequency. The focus at the feed point ensures effective power transfer from the feed line to the radiating elements, while the distribution along the slots and edges enhances the desired radiation properties.

At 5.76 GHz: The surface current concentrates more around the antenna edges and apertures. This shift corresponds to a higher-order resonant mode, where the shorter wavelength interacts more

prominently with finer structural details of the antenna. The higher frequency alters current propagation, emphasizing areas that enhance radiation efficiency at this frequency.

Antenna Efficiency through Surface-current Distribution: The efficiency of the antenna correlates closely with its surface current distribution. Analysis of current distribution at 2.45 GHz and 5.76 GHz indicates that in both cases, current concentration in specific regions facilitates efficient conversion of input power into radiated electromagnetic waves. The distinct current paths at these frequencies ensure effective operation across the dual-band spectrum.

Moreover, the observed current patterns suggest well-matched impedance at the feed point for both frequencies. This characteristic is critical for minimizing reflections and maximizing power transfer, thereby enhancing the overall antenna efficiency.

In summary, the simulated surface-current distribution underscores the antenna's capability to operate efficiently across its dual-band frequencies. At 2.45 GHz, concentration around the feed point and slots facilitates effective radiation, while at 5.76 GHz, focus near the edges and apertures supports higher frequency performance. These visions confirm the antenna's design effectiveness in achieving dual-band operation with high efficiency.

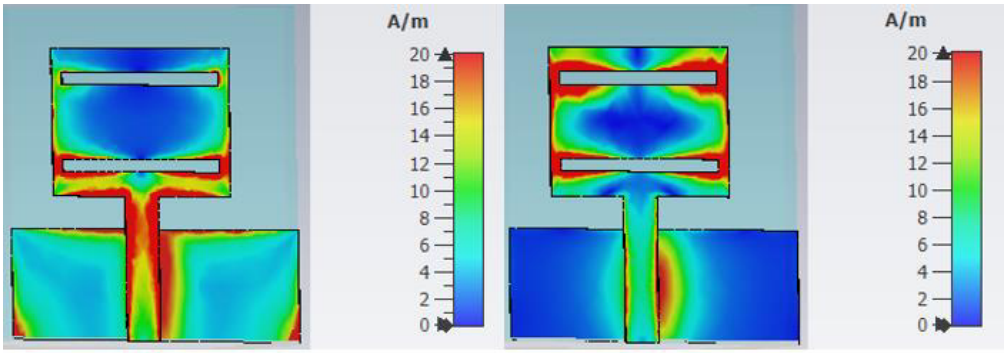


Figure 8. The surface-current distribution of the designed antenna at (a) 2.45 GHz and (b) 5.76 GHz.

4.2 Gain and Efficiency

Figure 9 presents the measured and simulated gain and efficiency of the antenna. At the 2.45 GHz frequency band, the gain is 2.65 dB, while at the 5.76 GHz resonance band, it is approximately 2.8 4 dB. The radiation efficiency at 2.45 GHz is about 95%, demonstrating the antenna's effectiveness in converting input power into radiated energy with minimal losses. This high efficiency is attributed to the optimized design and the incorporation of a Defected Microstrip Structure (DMS), which enhances impedance matching and reduces unwanted radiation and dielectric losses. At 5.76 GHz, the radiation efficiency is around 73%. This efficiency is satisfactory for high-frequency operation, where losses in dielectric and conductive materials are greater than at lower frequencies. Nonetheless, the antenna maintains a commendable level of efficiency across both operating bands. Figure 10 shows the 3D gain and 3D directional radiation pattern.

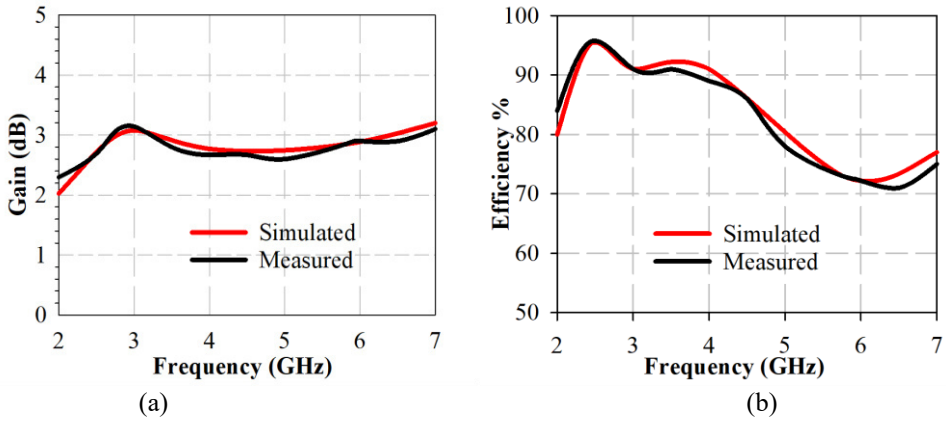


Figure 9. Measured and simulated (a) Peak gain, (b) Radiation efficiency.

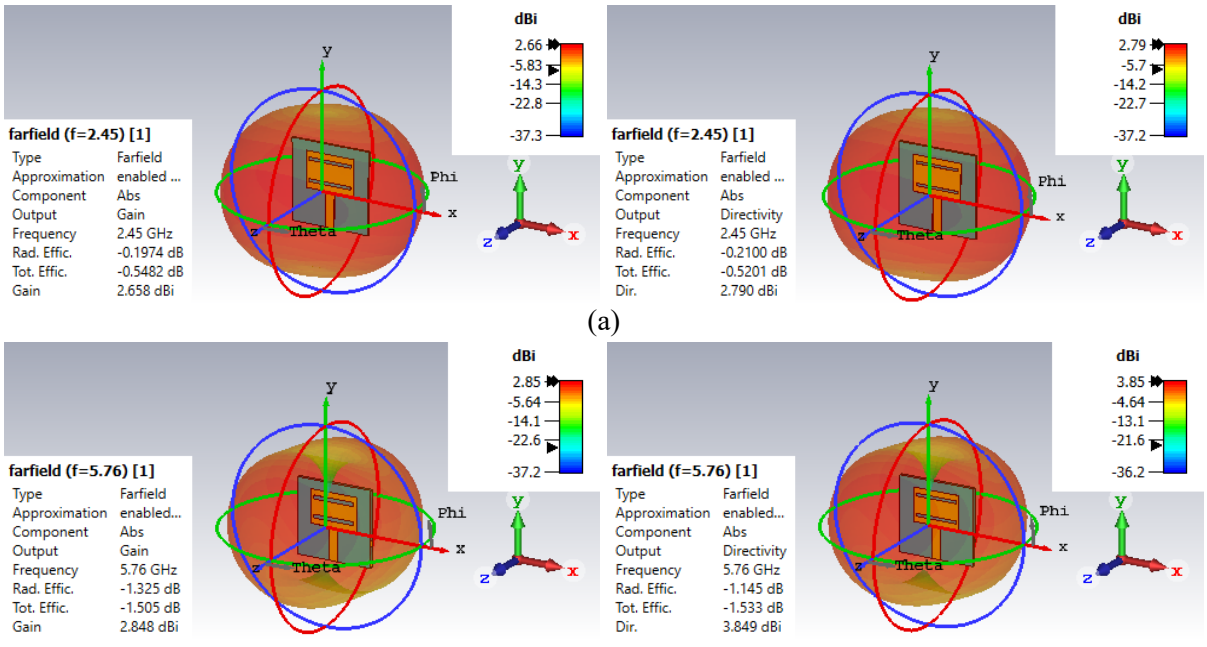


Figure 10. Gain and directivity 3D at (a) 2.45 GHz and (b) 5.76 GHz.

4.3 Radiation Pattern

The measured and simulated radiation patterns of the designed antenna in the E-plane and H-plane are shown in Figure 11. At the lower band (2.45 GHz), depicted in Figure 11(a), the radiation pattern in the E-plane is nearly omnidirectional, while in the H-plane, it is nearly bidirectional. At the upper band (5.76 GHz), shown in Figure 11(b), the antenna exhibits a nearly omnidirectional pattern in the E-plane and a directional pattern in the H-plane. Figure 12 illustrates the Geozondas antenna measurement system that was used to assess the antenna's radiation pattern. A horn antenna was used as the transmitting antenna, while the antenna under examination served as the receiving antenna.

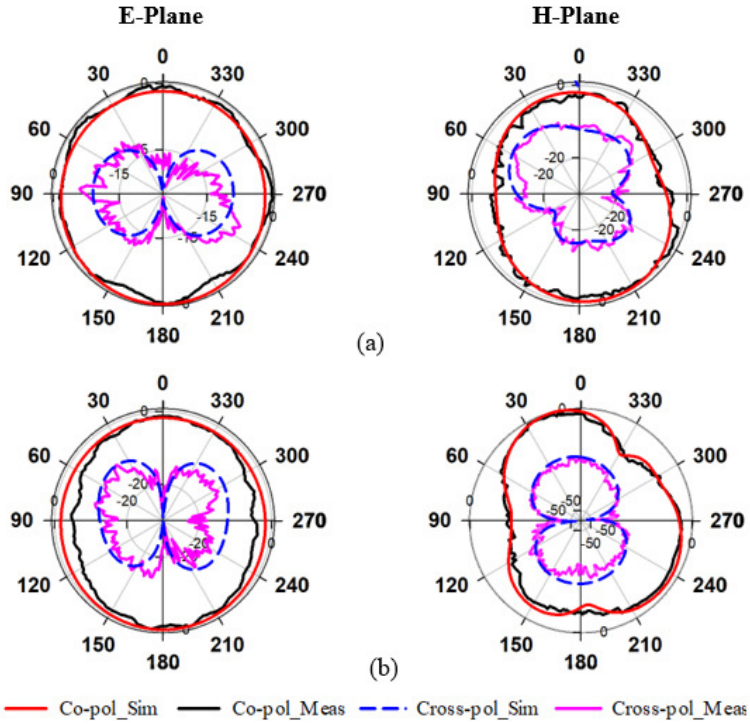


Figure 11. Radiation pattern of the designed antenna in the E-plane and H-plane at (a) 2.45 GHz and (b) 5.76 GHz.

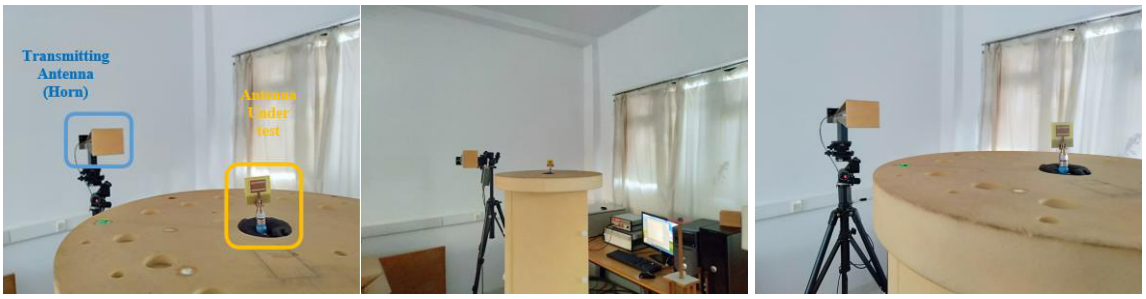


Figure 12. Measurement setup of the designed antenna.

5. PERFORMANCE COMPARISON

A performance comparison has been carried out between the designed antenna and recently reported antennas in the literature in terms of resonant frequencies, size, bandwidth, gain and efficiency, as shown in Table 2. It is observable that the designed antenna has a compact size compared to all the antennas proposed in the references listed in Table 2. In addition, the suggested antenna is distinguished by its elevated efficiency.

Table 2. Comparison of the antenna in this study with antennas proposed in previous studies.

Ref.	Frequency f1/f2 (GHz)	Dimensions (mm^3) Substrate Material	Bandwidth (GHz)	Peaks Gain (dB)	Efficiency (%)
2	2.4/5.2	40 × 30 × 0.8 FR-4	(2.27–2.58)12.7% (4.92–5.49)10.95%	… 	89 87
12	2.54/5.24	40×40×1.524 RogersTMM4	(2.24–2.93)26.69% (4.48–5.54)21.16%	2.5 2.7	…
18	2.4/5.8	47.3 × 55×1.6 FR-4	(2.3–2.492) 7.4% (5.586–6.06) 8.17%	2.5 2.8	…
21	2.4/5.25	52 × 60× 1.6 FR-4	4.2% and 2.3%	2 4.6	…
22	2.45/3.53	59.5×47× 1.6 FR-4	(2.43–2.49)2.44% (3.50–3.56)1.7%	2.45 3.53	…
23	2.45/5.8	70 ×70× 31 3D printed PLA	(2.39–2.52)5.29% (5.76–5.95)3.25%	7.8 6.8	90 80
24	2.44/5.5	52 ×40× 2 FR-4	(2.33–2.86)20.6% (5.76–5.95)15.7%	3.5 3.53	---
This work	2.45/5.76	27 × 27×1.6 FR-4	(2.13–2.88)29.94% (5.65–6.01)6.2%	2.65 2.84	95 73

6. CONCLUSION

This research presents an innovative dual-band patch antenna design with compact dimensions, tailored to meet the requirements of various modern wireless communication systems, including WLAN, ISM, Bluetooth, WiMAX and WiFi-2.4. The antenna operates efficiently at frequencies of 2.45 GHz and 5.76 GHz, achieving high efficiencies of over 95% and 73%, in both operating bands. Additionally, the antenna attains a gain of 2.65 dB at 2.45 GHz and 2.84 dB at 5.76 GHz. The antenna was manufactured and measured, the results being in good agreement with simulations. The measured -10 dB S11 bandwidths cover 750 MHz (2.13-2.88 GHz) and 330 MHz (5.65-6.01 GHz). This ensures that the antenna meets the frequency requirements for WLAN in both the lower and upper bands, ISM (2.4–2.5 GHz), WiMAX rel 1 (2.3–2.4 GHz), WiMAX rel 1.5 (2.5–2.69 GHz) and Bluetooth (2.407–

2.484 GHz). Based on these results and the antenna's characteristics, such as compact size and ease of manufacture, this design is suitable and highly practical for advancing antenna technology and enhancing wireless-communication systems.

REFERENCES

[1] L. C. Paul et al., "Wideband Inset Fed Slotted Patch Microstrip Antenna for ISM Band Applications," Proc. of the 2019 Joint 8th Int. Conf. on Informatics, Electronics & Vision (ICIEV) and 2019 3rd Int. Conf. on Imaging, Vision & Pattern Recognition (icIVPR), pp. 79–84, DOI: 10.1109/ICIEV.2019.8858553, May 2019.

[2] P. B. Nayak, R. Endluri, S. Verma and P. Kumar, "A Novel Compact Dual-band Antenna Design for WLAN Applications," arXiv, DOI: 10.48550/arXiv.2106.13232, May 12, 2021.

[3] P. B. Nayak, S. Verma and P. Kumar, "A Novel Compact Tri-band Antenna Design for WiMAX, WLAN and Bluetooth Applications," Proc. of the 2014 20th National Conf. on Communications (NCC), pp. 1–6, DOI: 10.1109/NCC.2014.6811379, Feb. 2014.

[4] P. B. Nayak, R. Endluri, S. Verma and P. Kumar, "Compact Dual-band Antenna for WLAN Applications," Proc. of the 2013 IEEE 24th Annual Int. Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), pp. 1381–1385, DOI: 10.1109/PIMRC.2013.6666356, Sep. 2013.

[5] P. Nayak, M. Garetto and E. W. Knightly, "Modeling Multi-user WLANs under Closed-loop Traffic," IEEE/ACM Transactions on Networking, vol. 27, no. 2, pp. 763–776, Apr. 2019.

[6] A. Ghaffar, X. J. Li and B.-C. Seet, "Compact Dual-band Broadband Microstrip Antenna at 2.4 GHz and 5.2 GHz for WLAN Applications," Proc. of the 2018 IEEE Asia-Pacific Conf. on Antennas and Propagation (APCAP), pp. 198–199, DOI: 10.1109/APCAP.2018.8538297, Aug. 2018.

[7] X.-Q. Zhu, Y.-X. Guo and W. Wu, "A Novel Dual-band Antenna for Wireless Communication Applications," IEEE Antennas and Wireless Propagation Letters, vol. 15, pp. 516–519, DOI: 10.1109/LAWP.2015.2456039, 2016.

[8] S. Abulgasem et al., "Antenna Designs for CubeSats: A Review," IEEE Access, vol. 9, pp. 45289–45324, DOI: 10.1109/ACCESS.2021.3066632, 2021.

[9] M. Dube, Design and Fabrication of a Miniaturised Dual Band Planar Antenna for Wireless Communication, PhD Thesis, University of Johannesburg, 2023. [Online], Available: https://ujcontent.uj.ac.za/esploro/outputs/graduate/Design-and-fabrication-of-a-miniaturised/9934308707691, Accessed: Jun. 30, 2024.

[10] M. M. Alam, R. Azim, N. M. Sobahi, A. I. Khan and M. T. Islam, "A Dual-band CPW-fed Miniature Planar Antenna for S-, C-, WiMAX, WLAN, UWB and X-band Applications," Scientific Reports, vol. 12, no. 1, Article no. 1, DOI: 10.1038/s41598-022-11679-7, May 2022.

[11] J. Ali, S. Abdulkareem, A. Hammoodi, A. Salim, M. Yassen and H. Al-Rizzo, "Cantor Fractal-based Printed Slot Antenna for Dual-band Wireless Applications," Int. Journal of Microwave and Wireless Technologies, vol. 8, no. 2, pp. 263–270, DOI: 10.1017/S1759078714001469, Mar. 2016.

[12] M. T. Yassen, M. R. Hussan, H. A. Hammas, H. Al-Saedi and J. K. Ali, "A Dual-band Printed Antenna Design Based on Annular Koch Snowflake Slot Structure," Wireless Pers. Commun., vol. 104, pp. 649–662, DOI: 10.1007/s11277-018-6039-0, 2019.

[13] G. Rushingabigwi, L. Sun, Q. Zhu, Y. Xia and Y. Yu, "Design and Realization of a Dual Wide Band Printed Monopole Antenna for WiFi and WiMAX Systems," IJCNS, vol. 09, no. 05, pp. 184–197, DOI: 10.4236/ijcns.2016.95018, 2016.

[14] R. Roshan, S. Prajapati, H. Tiwari and G. Govind, "A Dual Wideband Monopole Antenna for GSM/UMTS/LTE/WiFi/and Lower UWB Application," Proc. of the 2018 3rd Int. Conf. on Microwave and Photonics (ICMAP), pp. 1–2, DOI: 10.1109/ICMAP.2018.8354635, Feb. 2018.

[15] B. P. Nadh, B. T. P. Madhav and M. S. Kumar, "Design and Analysis of Dual Band Implantable DGS Antenna for Medical Applications," Sādhanā, vol. 44, no. 6, p. 131, DOI: 10.1007/s12046-019-1099-8, May 2019.

[16] H. Yao, X. Liu, H. Zhu, H. Li, G. Dong and K. Bi, "Dual-band Microstrip Antenna Based on Polarization Conversion Metasurface Structure," Frontiers in Physics, vol. 8, 2020. [Online], Available: https://www.frontiersin.org/articles/10.3389/fphy.2020.00279, Accessed: Oct. 30, 2023.

[17] L. C. Paul, H. K. Saha, T. Rani, R. Azim, M. T. Islam and M. Samsuzzaman, "A Dual-band Semi-circular Patch Antenna for WiMAX and WiFi-5/6 Applications," Int. Journal of Communication Systems, vol. 36, no. 1, p. e5357, DOI: 10.1002/dac.5357, 2023.

[18] R. H. Thaher and Z. S. Jamil, "Design of Dual Band Microstrip Antenna for Wi-Fi and WiMax Applications," TELKOMNIKA (Telecommunication Computing Electronics and Control), vol. 16, no. 6, Article no. 6, DOI: 10.12928/telkomnika. v16i6.10016, Dec. 2018.

[19] A. A. Yassin, R. A. Saeed and R. A. Mokhtar, "Dual-band Microstrip Patch Antenna Design Using C-slot for WiFi and WiMax Applications," Proc. of the 2014 Int. Conf. on Computer and Communication Engineering, pp. 228–231, DOI: 10.1109/ICCCE.2014.72, Sep. 2014.

[20] C.A. Balanis, Antenna Theory Analysis and Design, 2nd Edn, J. Wiley & Sons, New York, 2016.
[21] S. Luo, Y. Li, Y. Xia, G. Yang, L. Sun and L. Zhao, "Mutual Coupling Reduction of a Dual-band Antenna Array Using Dual-frequency Metamaterial Structure," The Applied Computational Electromagnetics Society Journal (ACES), vol. 2019, pp. 403–410, Mar. 2019.
[22] I. Zahraoui et al., "A New Planar Multiband Antenna for GPS, ISM and WiMAX Applications," Int. Journal of Electrical and Computer Engineering (IJECE), vol. 7, no. 4, Article no. 4, DOI: 10.11591/ijece. v7i4. Pp 2018-2026, Aug. 2017.
[23] K. N. Olan-Nuñez and R. S. Murphy-Arteaga, "Dual-band Antenna on 3D-Printed Substrate for 2.4/5.8 GHz ISM-band Applications," Electronics, vol. 12, no. 11, p. 2368, 2023.
[24] X. Wu, X. Wen, J. Yang, S. Yang and J. Xu, "Metamaterial Structure Based Dual-band Antenna for WLAN," IEEE Photonics Journal, vol. 14, no. 2, pp. 1-5, 2022.

ملخص البحث:

تقدم هذه الوقة البحثية هوائياً ذا نطاقين ترددّيين مصمّماً لأنظمة الاتّصالات اللاسلكية. ويعمل الهوائي المقترح بفاعلية عند تردّد 2.45 جيجاهيرتز وتردّد 5.76 جيجاهيرتز، محقّقاً أداءً عالياً على الرّغم من صغر حجمه. وقدّ تمّ استخدام بنية الشّريط الميكروي المنحرف (DMS) للحصول على النّطاقين التّردّديّين المرغوبين وتحسين توافق الممانعات عند التّردّدات المطلوبة. وتمّ تجميع الهوائي المقترح على طبقة أساس من نوع FR-4 بأبعاد بلغت (27 ملم × 27 ملم × 1.6 ملم). وقد بلغت فعالية المحاكاة عند التّردّدين 2.45 جيجاهيرتز و 5.76 جيجاهيرتز 95% و 73% على الترتيب. أما كسْب الهوائي المقترح فبلغ أكبر من 2.65 ديسيبل وحوالي 2.84 ديسيبل عند النّطاقين التّردّديين المرغوبين على الترتيب.

والجدير بالذّكر أنّ الهوائي المقترح تمّت محاكاتُه وتصنيعه وفحْصه والتّحقُّق من خصائصه عملياً. أمّا فقْد الإرجاع المقاس عند الرّنين الأدنى والرّنين الأعلى فكان 29.94% مغطياً النّطاق التّردّدي من 2.13 جيجاهيرتز إلى 2.88 جيجاهيرتز، و 6.2% مغطّياً النّطاق التردّدي من 5.65 جيجاهيرتز إلى 6.01 جيجاهيرتز على التّرتيب. كذلك تمّت المقارنة بيْن نمط الإشعاع المقاس للهوائي المقترح ونمط الإشعاع للهوائي المقترح عن طريق المحاكاة، حيث اتّفقا بشكلٍ كبير.

ويتميز الهوائي المقترح ببساطة تصميمه وسهولة تصنيعه وصغر حجمه، مع عمله بشكلٍ جيّد عند التّردّدات الّتي صُمّم لها، الأمر الّذي يجعله ملائماً للاستخدام في أنظمة الاتّصالات اللاسلكية.

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 10, No. 04, December 2024.

PROCESSING TOOLS FOR CORPUS LINGUISTICS: A CASE STUDY ON ARABIC HISTORICAL CORPUS

Bassam Hammo[1] and Sane Yagi[2]

(Received: 10-May-2024, Revised: 2-Jul.-2024, Accepted: 29-Jul.-2024)

ABSTRACT

This paper explores the development, design and reconstruction of a Historical Arabic Corpus (HAC), which covers more than 1600 years of uninterrupted language use. The study emphasizes the technical aspects followed to enhance the system and provide a usable concordancer, along with simple experiments conducted on the corpus and the concordancer. Arabic has a rich literary and cultural heritage spanning thousands of years. The inclusion of digital resources and the advancement in natural language processing (NLP) technology have made Arabic historical corpora increasingly crucial for researchers and learners worldwide. By integrating HAC and its tools into Arabic language learning, learners can delve deeper into vocabulary and culture and gain valuable insights that improve their language skills and understanding of Arabic. This combination of human guidance and NLP technology makes learning an engaging and enjoyable experience, offering a dynamic and authentic way to master the Arabic language.

KEYWORDS

Historical Arabic corpus, Corpus tools, Concordancer, Learning Arabic, Data normalization, Semantic shifting.

1. INTRODUCTION

The Arabic language is recognized for its linguistic richness. It has a rich and extensive history and is one of the most widely spoken languages in the world. Literature and historical texts have been produced from the early Islamic period to the modern era, providing a valuable resource for researchers and language learners. Understanding and analyzing Arabic requires a comprehensive and systematic approach as a language deeply rooted in literature, religion and daily life. One vital tool for linguistic research in the Arabic language is the compilation of text corpora and dictionaries [1].

A language corpus (plural corpora) is a collection of texts systematically organized and annotated for linguistic analysis. Constructing an Arabic-language corpus provides a fundamental resource for linguistic, cultural and historical studies. The following are a few benefits of what a dedicated linguistic Arabic corpus can do.

1. Providing linguists and researchers with a vast and varied dataset, enabling in-depth analyses of language patterns, syntax and semantics. It also facilitates investigations into language evolution and usage across different regions.
2. Offering educators and language learners an extensive range of authentic materials representing the language in diverse contexts helps develop effective teaching methodologies, curriculum design and language-proficiency assessments.
3. Contributing to preserving cultural heritage and becoming a repository of cultural expressions, idioms, philosophies and social standards.
4. Providing an essential tool for developing applications and advancing research in areas like natural language processing (NLP), machine learning (ML), deep learning (DL), data mining (DM) and large language models (LLM). For instance, building robust language models, sentiment-analysis tools and machine translation systems relies on the availability of high-quality linguistic data.

Historical corpora, another type of text corpora, are extensive collections of written texts compiled and organized for educational and research purposes. They provide researchers and learners access to various historical documents, allowing them to study the evolution of a language and gain insights into a particular cultural and social history from different eras. These corpora contain many historical documents, such as religious texts, legal texts, scientific works and other genres. Arabic historical corpora are essential resources for researchers interested in studying the history of the Arab world and

1. B. Hammo is with the Department of Computer Information Systems, KASIT, The University of Jordan and with the Department of Software Engineering, Princess Sumaya University for Technology, Amman, Jordan. Email: b.hammo@ju.edu.jo, b.hammo@psut.edu.jo
2. S. Yagi is with the Department of Foreign Languages, University of Sharjah, United Arab Emirates. Email: syagi@sharjah.ac.ae

Islamic civilization.

In this study, building upon the groundwork described in [2]-[3], we continue the efforts to compile and refine a historical corpus of Arabic texts named HAC, spanning 1600 years of language evolution. The previous attempts focused on introducing the essential tools required for assembling HAC and implementing NLP techniques to preprocess the text, ultimately organizing it into a structured eXtensible Markup Language (XML) schema, allowing for search and analysis. The HAC corpus is a valuable resource for linguists and is designed for Arabic-language learners studying at the University of Jordan. It enables them to delve into the rich linguistic heritage of millions of textual instances.

The corpus is designed to serve multiple purposes, including supporting Arabic-language learners and advancing research in NLP, data mining and other computational fields. The extensive dataset and the developed tools are tailored to the needs of both educational and research-oriented users. Researchers in NLP and other fields benefit from the structured and annotated data for developing and testing their models. The need for a new concordancer tool is due to the limitations of existing tools when applied to historical Arabic texts. Many of these tools are not optimized for the Arabic language's unique morphological and orthographic complexities, particularly in a historical context. Existing tools often fail to handle search variations critical for accurate historical-corpora analysis. The new concordancer tool addresses these gaps by offering enhanced morphological analysis capabilities, adapting to historical-text variations and providing advanced customization options. This tool is essential for researchers and linguists to perform accurate searches and analyses on historical texts, making it a significant advancement in Arabic-corpus linguistics.

The remainder of this paper is structured as follows: Section 2 presents background information on corpus linguistics. Section 3 reviews pertinent prior research. Section 4 presents the methodology and technical processes that we have followed to reconstruct the HAC corpus. Section 5 showcases experiments demonstrating the practical utility of our efforts. Finally, Section 6 concludes the work and outlines future-research avenues.

2. BACKGROUND

2.1 A Brief Introduction to Corpora and Corpus Linguistics

Corpora encompass extensive and structured collections of texts from diverse sources, accommodating written texts such as books, articles, websites and transcriptions of spoken language. These electronically stored and processed repositories are often augmented with annotations and metadata, such as Part-of-speech (POS) tags and morphological information, to enhance their utility for specific research objectives [2]-[4]. Corpora are vital resources for linguistic and computational linguistic research and the study of language use in real-world contexts. They are fundamental to applications in lexicography [5], translation [6], language learning and teaching [7] and data mining [8].

Historical corpora are particularly significant for historical linguists. They comprise texts spanning the entire history of a language or specific eras. These corpora illuminate the evolution of languages over time, revealing shifts in word meanings and grammatical structures. They play a significant role in compiling historical dictionaries [1], offering insights into semantic changes and providing illustrative quotations for word senses.

Whether we are designing a contemporary or historical corpus, details regarding sources' authorship, publication date, genre and broader context should be documented to enrich the corpus with valuable insights that aid researchers in interpreting and categorizing linguistic data.

Corpora and their associated tools facilitate language understanding and learning, offering valuable resources for educators and learners [9]. These tools are essential in language acquisition, contributing to vocabulary expansion, contextual learning and a deeper understanding of grammar and syntax. For instance, adaptive learning systems can utilize corpus data to individualize learning experiences [10]. These platforms customize language learning based on individual strengths and weaknesses and optimize the learning process for each student. Other essential tools might include the following [11]:

Tools for the Creation of the HAC

1. Language-analysis software: We utilized a concordancer, Khoja's stemmer and Stanford part-of-

speech tagger to analyze the language in HAC, ensuring accurate annotation and tagging of the historical texts.
2. Metadata-management tools: Tools used to manage and catalog information about each text in the corpus, including author, publication date and title, facilitating easy search and retrieval.
3. Corpus-management software: We employed a database-management system to manage the HAC, allowing for efficient storage, retrieval and updates to the corpus data.
4. Computational tools for analysis: NLP tools were integral in analyzing the HAC, enabling tasks, such as tokenization and light stemming, which are crucial for handling the linguistic complexity of Arabic texts.

Tools for Exploitation of the HAC

1. Concordancer: Researchers and language students can use the concordancer to search and analyze specific words or phrases within HAC, helping understand historical usage patterns and contexts.
2. Visualization tool: Visualizes linguistic data to uncover patterns and trends within the HAC, which are valuable for linguistic and historical research.

2.2 The Historical Arabic Corpus (HAC)

The initial historical Arabic corpus (HAC) was initiated in 2015. It was constructed using a corpus-builder system developed in Java to compile and encode its data into an XML schema automatically. Interested readers who want to learn more about the corpus-builder system are referred to [2]-[3]. The input to the corpus builder was a text document encoded in UTF-8 with its meta-data [2]. We integrated a stemmer and a part-of-speech (POS) tagging modules in the corpus-builder system to build the final corpus. We adapted an Arabic stemmer developed by Khoja [12] to extract a root, stem and morphological pattern for each word. For POS tagging, we employed the Stanford Part-Of-Speech Tagger [13]. Both tools were popular when the project was started, providing essential functionalities required for processing Arabic text.

Since then, newer tools such as Farasa [14] and Computational Approaches to Modeling Language Lab (CAMeL) [15] have been developed, offering enhanced capabilities in Arabic NLP. These tools could be considered for future updates and improvements to HAC, potentially increasing the accuracy and efficiency of text processing.

Project Goals and Implementation

The primary goal of the HAC project is to create a comprehensive, searchable database of historical Arabic texts. This involved several key objectives:

1. Elucidating the conceptual and technical refinements applied to the HAC corpus, including the normalization procedures employed to enhance its consistency and coherence.
2. Outlining the technical methodologies employed in constructing a database and searchable indices, incorporating various simplified and normalized tokens to facilitate efficient information retrieval.
3. Presenting the design of a new concordancer tool developed with user-friendly interfaces, providing researchers with a platform to experiment with and analyze the corpus.
4. Experimenting with the HAC corpus and the enhanced concordancer.

To achieve these goals, the following steps were undertaken:

1. A robust database architecture was designed to facilitate efficient storage and retrieval of text data. This included restructuring the data-storage system to handle the large volume of text and metadata.
2. Advanced search algorithms were implemented to enable precise and fast data retrieval. This included the development of custom-search interfaces tailored to the needs of researchers.

2.3 The Concordancer

A concordancer is a software tool used in linguistics and language analysis to identify and analyze the frequency, distribution and usage of words and phrases in a text corpus. It helps determine the context in which a word or phrase appears and displays the lines of text containing the word or phrase and its neighboring words. This enables researchers to study how words and phrases are used in different contexts and how they are related. The following points highlight a few of the benefits of using a concordancer.

1. Teachers can utilize a concordancer in language teaching to help students understand the usage of words and phrases in contexts and to develop their vocabulary and grammar skills [16]-[19].
2. Translators can use a concordancer to identify the most appropriate translation of a word or phrase in a given context [20]-[21].
3. Linguists can use a concordancer to study language use and patterns, such as the distribution of words and phrases across different genres, periods or social groups [22]-[23].
4. Researchers in fields, such as literature, history and sociology, can use a concordancer to analyze text to identify patterns and trends in the data [24].

2.4 Initial Challenges and Limitations

At the early stage of constructing HAC, a few problems were raised. The major problem was associated with the scalability of data storage, while the second one was related to searching and retrieving the data effectively. The primary users of the system were students from the Linguistic Department at the University of Jordan. They know little about computers. As the volume of data started growing, the response time of inquiring data from the XML database turned out to be very slow and the system's GUI was not user-friendly. Therefore, it was obvious that the original data structure and the GUI design were unsophisticated and needed to be revised and enhanced. This study aims to solve this problem by replacing the XML schema with a sophisticated relational database-management system running on the server side with optimized queries and a friendlier concordancer system.

Another issue that we are still striving to solve is that HAC needs to be balanced and requires more historical Arabic text in digital format, which makes it, in its current state, unrepresentative of the genres and eras that it should cover [2]. As part of our continuous-improvement efforts, we added further five million terms to HAC, bringing the total number of terms to 50 million terms.

3. LITERATURE REVIEW

The development of Arabic corpora is still in its early stages [25]. Initially, Arabic corpora were mainly created through manual efforts or basic tools that compile texts into XML format, often accompanied by metadata annotations [26]. The UAM Corpus Tool, developed by the Universidad Autónoma de Madrid [27], is a comprehensive software suitable for corpus linguistics research. It offers functionalities for corpus compilation, annotation and analysis, including concordancing, collocation analysis and statistical-processing tools. It utilized XML as its underlying data-storage format and facilitated cross-layer searching, semi-automatic tagging, statistical reporting and visualization of tagged data.

Later, plenty of contemporary Arabic corpora were designed with a range of structures and annotations [28]-[29]. Another avenue encompassed Quranic and Hadith (Prophet Mohammad's traditions) corpora [30]-[33]. Other scholars focused on designing tools for the Arabic language, such as the work of [34], where the authors proposed iSPEDAL, an enhanced electronic dictionary for the Arabic language. A corpus and a set of tools to experiment with contemporary Arabic were introduced in [35]. The corpus included editorials of newspapers collected from different countries, Arab countries' constitutions, dictionaries and the Holy Quran, in addition to news from sports, technology and politics.

The development of Arabic corpora has faced challenges, but significant progress has been made. Resources such as the Multilingual Annotated Standard Dataset of Educational Resources (MASADER) [36] and the Linguistic Data Consortium (LDC) catalog [37] have played a vital role in advancing Arabic corpus linguistics. MASADER provides a comprehensive catalog of Arabic-language resources, including datasets and tools for various NLP tasks. The LDC catalog includes extensive Arabic-language resources such as speech and text corpora, lexicons and annotated linguistic data. Building on these foundations, the HAC corpus aims to provide additional resources and tools tailored to historical Arabic texts.

English historical corpora, comprising samples of texts from earlier eras, are instrumental in studying language variation, changes and development. Examples of well-known English historical corpora accessible through the web are given in Table 1. For instance, the Helsinki Corpus of English Texts is a structured multi-genre corpus spanning Old, Middle and Early Modern English periods, offering insights into linguistic forms, structures and lexemes across different epochs. Similarly, ARCHER (A Representative Corpus of Historical English Registers) presents a multi-genre corpus of British and American English, covering the period from 1600 to 1999.

Table 1. Examples of well-known projects encompassing English historical corpora.

English historical corpora	**URL address (*Accessed on April 20, 2024*)**
The British National Corpus (BNC))	http://www.natcorp.ox.ac.uk/corpus/
The Penn Treebank (PTB)	https://catalog.ldc.upenn.edu/LDC99T42
Helsinki Corpus of English Texts	https://varieng.helsinki.fi/CoRD/corpora/HelsinkiCorpus/
ARCHER: A Representative Corpus of Historical English Registers	https://www.projects.alc.manchester.ac.uk/archer/

For Arabic historical corpora, the most relevant examples are the King Saud University Corpus of Classical Arabic (KSUCCA) and the King Abdul-Aziz City for Science and Technology (KACST) Arabic corpus [2]. KSUCCA, although supposedly encompassing classical Arabic texts from the pre-Islamic era until 1100 C.E., it lacks comprehensive coverage and evidence of representativeness [38]. Similarly, while the KACST Arabic corpus aims to be a comprehensive resource spanning various periods and domains, efforts are ongoing to enhance its representativeness and balance [39].

Other examples of well-known Arabic historical corpora projects accessible through the web are given in Table 2. The projects highlight the growing efforts to digitize, preserve and make accessible Arabic historical datasets, enabling researchers to investigate the rich history of the Arabic world. These projects are just a few examples of the wide range of ongoing research on Arabic historical corpora and there may be additional recent works since our knowledge cut-off date in December 2023.

Concordancing tools are crucial in linguistic analysis, aiding language learners and researchers in vocabulary acquisition, collocation identification and grammatical comprehension. While English boasts numerous concordancing tools, Arabic offerings are relatively limited. Earlier works included AntConc [40], aConCorde [41], AraConc [42].

While these tools serve essential functions, there remains a need for further research to develop sophisticated schemas accompanied by tools tailored to handle morphological annotation and facilitate automated Arabic-corpora construction. Moreover, advancement in Arabic NLP is slower than in English due to a scarcity of freely available corpora, lexicons and sophisticated machine-readable dictionaries, underscoring the need for concerted efforts to advance research in this area.

Table 2. Examples of well-known projects encompassing Arabic historical corpora.

Arabic historical corpora	**URL address (*Accessed on April 20, 2024*)**
The Digital Library of the Middle East (DLME) from Stanford Libraries represents a platform combining data collections from various cultural heritage institutions worldwide. It offers free and open access to the rich cultural legacy of the Middle East and North Africa.	https://dlmenetwork.org/library
The Qatar Digital Library (QDL) is a massive online repository that offers access to a diverse collection of historical documents related to the Gulf and Middle East. The collection includes manuscripts, maps, photographs and archival materials that provide insights into the Arabic world's social, cultural and political history.	https://www.qdl.qa/en
Al-Maktaba al-Shamela is a digital library that hosts a vast collection of classical Arabic texts, including religious, historical, literary and scientific data. It offers a comprehensive platform for accessing and searching thousands of Arabic manuscripts and books [43].	https://shamela.ws/
The King Saud University Corpus of Classical Arabic (***KSUCCA***) is a 50 million tokens annotated corpus of Classical Arabic texts from the period of pre-Islamic era (7^{th} Century CE) until the fourth Hijri century (11^{th} Century).	https://sourceforge.net/projects/ksucca-corpus/

This study aims to refine the HAC corpus through various technical improvements, including using normalization procedures to enhance consistency and coherence, the creation of a database with optimized queries for efficient information retrieval and the development of a user-friendly concordancer tool. By accomplishing these objectives, this research provides a solution to current challenges and contributes to the progress of Arabic-corpus linguistics and language-learning methodologies.

The refined HAC corpus and its accompanying tools will be a valuable resource for linguists, researchers, educators and learners, enabling them to conduct detailed analyses of Arabic-language patterns, syntax and semantics and provide authentic materials for language acquisition. This research represents a significant step towards bridging the gap between historical Arabic corpora and present-

day language-learning needs, delivering a dynamic platform for studying and mastering the Arabic language.

Studying authentic historical Arabic texts can provide valuable insights into classical Arabic grammar, vocabulary and stylistic conventions not commonly found in modern language-learning materials. While primarily beneficial for researchers and linguists, historians and those interested in digitizing Arabic cultural heritage can also benefit from the HAC corpus. However, it may have limited direct application for beginners learning Arabic as a foreign language. The HAC corpus primarily serves the needs of historians, linguists, researchers of the Arabic language and those interested in digitizing Arabic cultural heritage. Here is how each party may benefit from HAC:

1. HAC aims to preserve and provide access to historical Arabic texts, which are valuable for researchers studying the evolution of the Arabic language, linguistic variations over time and historical events documented in Arabic sources.
2. Linguists can use the corpus to analyze language usage, semantic changes and syntactic structures in historical contexts, aiding in understanding how the language has evolved and adapted across different periods of history.
3. Digitizing historical Arabic texts preserves cultural heritage and promotes awareness and appreciation of Arabic literature and history.

Differentiation from Other Arabic Historical Corpora

1. Scope and coverage: The HAC corpus has eight genres and around 50 million words distributed among predefined eras spanning 1600 years from the pre-Islamic era to the twenty-first century.
2. Accessibility and tools: An extensive dataset and a concordancer tailored to the needs of both educational and research-oriented users. Researchers in NLP and other fields benefit from the structured and annotated data for developing and testing their models.

4. RESEARCH METHODOLOGY

The methodology we employed in this study is depicted in Figure 1. It incorporates four stages: (1) data collection, (2) data pre-processing, (3) constructing the HAC database and indices and (4) experimentation with HAC and the concordancer. In the following subsections, we discuss each stage in more detail.

4.1 Data Collection

The HAC corpus was planned to include all primary classical Arabic text material available online, using the automated tools described in [2]. We searched the internet because of the limited resources for free digitized Arabic text and found Al-Maktaba al-Shamela, a free and open-source digital library (available at https://shamela.ws/) [43]. It has a wide range of religious, historical, literary and scientific data, making it an excellent platform for accessing and searching thousands of Arabic manuscripts and books.

To assemble a comprehensive corpus, we carefully considered the issue of textual representation before collecting the corpus data. Our primary concern was that the collection should cover all periods of Arabic history as recommended by Arabic literary historians [44]. This approach helps us understand the development and evolution of Arabic literature over time.

From previous work, we collected over 50 million tokens [2]-[3]. We classified the data into different eras every 100 years, beginning from Classical Arabic (pre-Islamic times) and ending with Modern Standard Arabic (MSA) of the current century. In addition, the corpus data was categorized into primary and secondary sources based on their representation of the language used during the time of authorship [2]. Primary texts, such as poetry, literary prose and non-fiction, offer insight into contemporary language practices without commenting on older texts. In contrast, secondary texts, like Quran exegesis and critical analyses of ancient poetry, provide commentary reflecting the language usage of the commentator's era, shedding light on linguistic customs from earlier times.

Similar to the works [2]-[3], genres conventionally influence the language used in a text. They are considered significant factors when representing texts. Consequently, we categorized the texts into eight genres: Dictionaries, Literary Prose, Poetry, History, Philosophy, Religion, Science and Thought. Apart

from the era, genre and primary/secondary categorization, we gathered general information about the texts, such as document title and author, to compile them into the corpus. We are also working on another text distribution based on regions and varieties of Arabic dialects, which we plan to include in future work. Table 3 shows various text examples from the HAC corpus based on historical principles and annotations regarding genre, author and era.

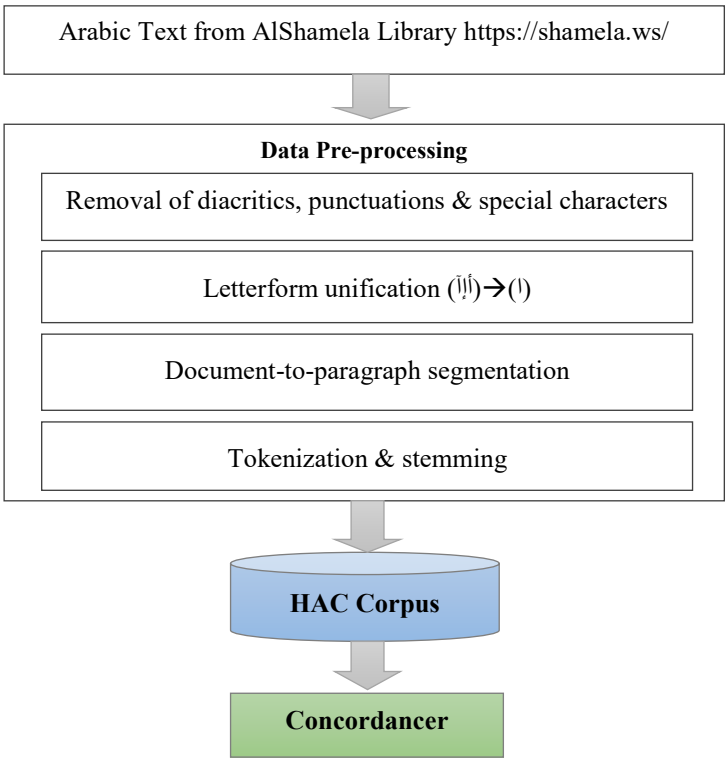


Figure 1. Methodology flow diagram.

Table 3. A sample of HAC data resources (collected from https://shamela.ws/).

	Title	Author	Era
Dictionaries	العين	الفراهيدي	700-800
	الصحاح تاج اللغة وصحاح العربية	الجوهري	900-1000
	مفردات القرآن	الأصفهاني	1100-1200
	مختار الصحاح	الرازي	1200-1300
	القاموس المحيط	الفيروزابادي	1400-1500
	تاج العروس 3-1	مرتضى الزبيدي	1700-1800
Poetry	ديوان امرئ القيس	امرؤ القيس	< 600
	شعر زهير بن أبي سلمى	زهير بن أبي سلمى	< 600
	ديوان حسان بن ثابت	حسان بن ثابت	600-700
	ديوان كزهر اللوز أو أبعد	محمود درويش	1900-2000
	الأعمال الشعرية الكاملة لابراهيم طوقان	ابراهيم طوقان	1900-2000
Philosophy	تهافت الفلاسفة	الغزالي	1000-1100
	تلخيص الخطابة	ابن رشد	1100-1200
	حي بن يقظان	ابن طفيل	1100-1200
	المختصر في المنطق	محمد بن محمد ابن عرفة	1300-1400
	أهل المدينة الفاضلة	الفارابي	1500-1600
	تاريخ الفلسفة الحديثة	يوسف مكرم	> 2000
Religion	القرآن الكريم	كلام الله عز وجل	600-700
	صحيح البخاري	أحاديث الرسول – البخاري	600-700
	الموطأ	الإمام مالك	700-800
	سنن أبي داوود	أبو داوود	800-900
	روضة العقلاء	ابن حبان	900-1000
	الأذكار	النووي	1200-1300
	تفسير الجلالين	جلال الدين المحلي و جلال الدين السيوطي	1400-1500
	شرح مسند أبي حنيفة	الامام القاري	1500-1600

4.2 Data Pre-processing

Arabic is a highly derivational and inflectional language. To handle the different ways in which Arabic text can be represented, we applied several normalization techniques described in the works [45]-[47]. These techniques utilize the indices for efficient search through the database, while the content of the texts in the database should be preserved to maintain its originality and integrity.

Data pre-processing starts with converting a text document D_i into the UTF-8 universal encoding, which represents every character in the Unicode character set, including Arabic characters. Further, a set of tasks is applied to extract the following information for each word w_j in D_i: word's root, pattern, part of speech and stem, but stop words were not removed.

It is noteworthy to mention that in our approach, we strictly maintain the integrity of the original text. To improve search capabilities, we applied pre-processing steps, such as tokenization, stemming and root extraction to parallel text versions, not the original. These parallel versions were used exclusively for search and retrieval, ensuring that the original text remains unaltered and can be accessed in its original form. This methodology allows us to provide efficient search functionality while preserving the authenticity of the historical documents.

The pre-processing steps were handled automatically and included the following tasks:

1. ***Normalization*:** Building a corpus requires normalization before exploring its content. Arabic-text normalization usually involves removing punctuation, stripping numbers out, removing diacritical marks, …etc. Root extraction, for example, is essential for effective searching and frequency-based analysis, so that words such as (كاتب, *writer*), (مكتبة, *library*) and (مكتب, *office*) can all be correlated to the third person singular root (كتب, *he writes*). Sometimes, normalization also includes stemming words, so that words such as (الكاتبون, *writers*), (والكاتبون and *writers*) and (الكاتبات, "*female*" *writers*) can all be stemmed to (كاتب, *writer*) and hence are not considered different words, as they all represent the same concept. Indices might also be normalized to prepare a textual database for searching. For instance, if a search for (لَعِبَ, *he played*) is intended to match the words (لَعِبْ, *play*) and (لُعَبْ, *toys*), then the text would be normalized by removing the diacritical marks to be all represented by one token (لعب). A set of steps is applied to reduce the number of extracted terms. They include:
   a) The removal of nonletters and special characters.
   b) The removal of non-Arabic letters.
   c) The replacement of initial أ،إ،آ with bare alef ا.
   d) The replacement of knotted ة (*ta marbuta*) with ه *(ha)*.
   e) The replacement of ending dotless yeh (alef maksura, ى) with yeh ي.
   f) The removal of leading proclitic particles, such as definite article, prepositions and conjunctions, trailing haa (ه), trailing Yeh_Yeh_Noon (يين), trailing Waw_Noon (ون), trailing Haa_Alef (ها), trailing pronominal enclitics used for dual and masculine plural forms (هم، هما).
   g) The removal of single-tone letters, such as Waw (و) and those produced by the above normalization steps.
2. ***Splitting documents into paragraphs*:** This process breaks a text document D_i into n paragraphs at the boundaries of paragraphs.
3. ***Tokenization*:** This process analyzes the paragraphs and splits them into individual token (word) streams. The boundaries of words, such as whitespaces and punctuation marks, are determined in this process.
4. ***Stemming and root extraction:*** A shallow stemming approach was applied to remove common affixes (i.e., prefixes and suffixes) from each word to extract its stem. This helps simplify words for frequency-based analysis and searching. For example, the words (الكاتبون, writers), (والكاتبون and writers) and (الكاتبات, female writers) would be stemmed to (كاتب, writer). Meanwhile, Khoja's algorithm was utilized to extract the roots of words. This process involves a deeper morphological analysis to identify the core set of letters that convey the fundamental meaning of the word. As an example: The words (كاتب, writer), (مكتبة, library) and (مكتب, office) would all be correlated to the root (كتب, write).

   In this study, we addressed ambiguity across clitics and stems, such as 'وجد,' which can mean "he

found" or "and grandfather," using a simple approach based on basic linguistic rules and lexical analysis applied in the stemmer. Due to resource limitations when the project began in 2015, we did not implement advanced techniques, such as contextual analysis or machine learning, for disambiguation. This limitation requires further investigation in future studies.

5. ***Part-of-speech tagging***: The process of assigning a part of speech to each word in a sentence, such as a noun, verb, adjective and more. It is crucial in determining how sentences are constructed from smaller units. POS tagging is widely used in syntactic and semantic analysis of sentences. We used the Stanford tagger, an open-source package written in Java programming language, to assign part of speech tags to words [10]. The package includes two trained tagger models for English and tagger models for Arabic, Chinese, French and German.

4.3 Motivation for New Data Storage

When designing HAC, we considered a portable and adaptable storage structure that is readable by humans and computers. Therefore, an XML schema was developed, including metadata tags for each document and annotation text for each token's morphology. Each token was stored in a single tag, along with its annotation attributes, such as the root, morphological pattern, POS tag and stem [2]-[3]. However, as the XML corpus grows, searching becomes slower, but its accessibility by text editors and portability make it appealing. Interested readers could refer to [2] to learn about the corpus structure.

We switched from utilizing the XML schema to a database to manage and manipulate the data in HAC. The change is because databases ensure scalability to handle large volumes of data with high levels of integrity and reliability and offer structured storage with the capability to define tables, relationships and constraints. In addition, databases provide powerful query capabilities, allowing for efficient data retrieval. We redesigned the database using the MS SQL-Server database-management system running on the server side to make it accessible through the web. For a more productive search, we ended up with three relations, as shown in Figure 2. They are as follows:

1. ***The Genre Table***: Stores information about each document *D* in the corpus and has the following attributes:
   - *docid*: a unique primary key assigned to each document composed of a genre, era and sequence number.
   - *path*: the actual path to the document.
   - *title*: document's title.
   - *author*: document's author.
   - *year*: document's year.
   - *era*: the era to which the document belongs
   - *category*: stores one of two values: primary or secondary.
   - *region* & *variety*: to be used in the future to store the region of the document and the dialect of that period.
2. ***The Paragraph Table***: Stores the document content after being split into paragraphs where a record has the following attributes:
   - *docid*: the document *id*.
   - *lineid*: a number assigned to a paragraph extracted from each document *D*.
   - *context*: the actual text of a paragraph.
3. ***The Posting Table***: Stores and tracks the occurrences of words, roots, …etc., associated with each text line in a document *D*. It has the following attributes:
   - *postid*: a unique number assigned to each post.
   - *docid*: the document ID from where the posting originated.
   - *lineid*: the number of the paragraph from where the posting originated.
   - *word*: the word in a text line after being normalized as described earlier.
   - *stem*: the word's stem after being processed as described earlier.
   - *root*: the word's processed root.
   - *tag*: the part of speech tag assigned to the word within the context.
   - *pattern*: an annotation assigned to each word based on the position of the three consonants (فعل) and the affixes. Patterns help in understanding the meaning of words.
   - *TF*: term frequency. Counts the occurrences of a word within a document *D*.
   - *DF*: document frequency. Counts the occurrences of a word within the entire corpus. TF and

DF are used to assign a weight for each token in the corpus.

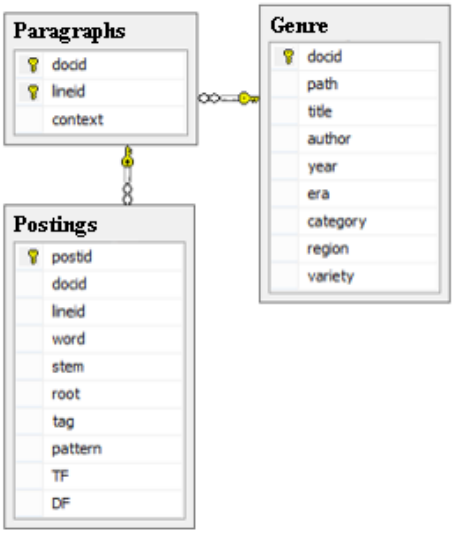


Figure 2. Database schema of the HAC corpus.

4.4 The Design of a New Concordancer

The primary purpose of a concordancer is to retrieve and display a text from the corpus in short contexts. It should allow the user to observe how a term is used within the context and how it might develop semantically within a period. Linguistic students at the University of Jordan heavily influenced HAC's concordancer. The initial design did not meet the users' satisfaction and a new, friendly design was in demand. The new concordancer includes alphabetical listings of all words in the historical corpus classified based on eras, genres and two main categories: primary or secondary. Searching through the concordancer shows where the terms (words, stems or roots) occur throughout all text.

The need for a new concordancer is emphasized by the limitations of existing tools when applied to historical Arabic texts. Many of these tools are not optimized for the Arabic language's unique morphological and orthographic complexities, particularly in a historical context. Existing tools lack the flexibility to analyze texts across different historical genres and eras. Our new concordancer is tailored specifically for the HAC corpus to address these gaps by offering enhanced morphological search capabilities, adapting to historical text variations and providing advanced customization options. This tool is essential for researchers and linguists to perform searches and analyses on historical texts, making it a significant advancement in Arabic-corpus linguistics. The concordancer provides the following functions:

1. Creating your word lists (vocabulary table) and producing concordances.
2. Searching for collocations and learning about a word's usage within neighboring words.
3. Counting word frequencies based on different eras and genres.
4. Discover a writer's stylistic traits by searching through authors.
5. Learning about all root derivatives and seeing each within the text's context.
6. Exploring results of searches to Excel sheets for further offline processing and analysis.

To illustrate the present interface of the concordancer, Figures 3 and 4 present screenshots from the new rendered version. Figure 3 shows the main components of the concordancer. The search selection tab has five options: word, stem, root, pattern and POS tag. The advanced tab allows a user to search for neighboring words around the word under search. If a search is performed on a root, a root-derivative list can be loaded to see all words derived from this root. A paragraph-up and paragraph-down offer an option to retrieve the previous and the following paragraphs for a selected paragraph from the grid. These functions are beneficial for linguists and researchers to understand and clarify the meaning of a word. Another search option is "Author," where a user can filter the search results by a particular author to learn about his/her writing style, for example. Finally, the statistics tab provides many HAC-related statistics and the user can export all his/her findings to an Excel sheet for further processing. Figure 4 shows a tracking of the Arabic root (عدل) meaning (modify, alter or adjust) in the literary prose genre in the era (700-800).

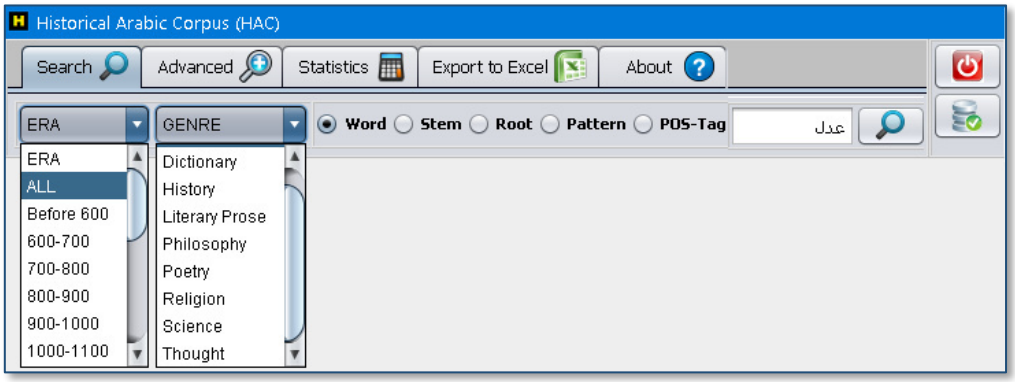


Figure 3. The concordancer and its functions.

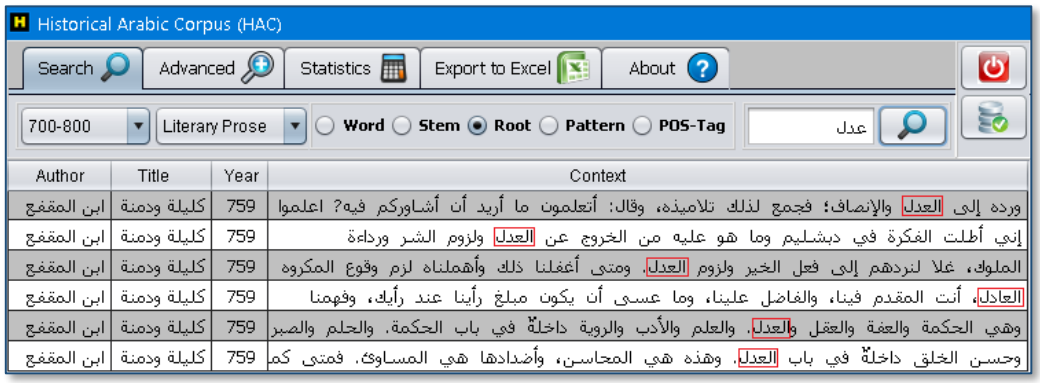


Figure 4. The concordancer tracking the Arabic root (عدل) in Literary Prose in the period 700-800.

5. Experiments and Discussion

5.1 Corpus Statistics

In this experiment, we analyzed HAC statistically to gain insights into the prominent words and their frequencies. The corpus comprises texts from various historical sources, providing a rich resource for understanding linguistic patterns and historical themes.

The HAC corpus has eight genres and around 50 million words, as depicted in Table 4. Table 5 shows the distribution of words among pre-defined eras spanning 1600 years from the pre-Islamic era to the twenty-first century. As one might notice from Table 4, most of the words (82%) fall under two genres: 49% in Literary Prose and 33% in History. Table 5 shows that 14% of the words were found in the (after 2000) era, while three eras have equal distributions, each representing 11%. Unfortunately, HAC is not yet balanced and still unrepresentative regarding genres and eras that it is supposed to accommodate. Our ambition is to create a representative and balanced corpus for the future.

Table 4. The Historical Arabic Corpus (HAC) (source: https://shamela.ws/).

Genre	Number of documents	Pct.	Number of paragraphs	Pct.	Number of words	Pct.	Number of distinct words	Pct.
Dictionaries	9	1.6%	183,059	5.2%	2,605,962	5.2%	320,484	13.2%
History	143	24.7%	1,163,937	33.0%	16,479,162	32.9%	642,901	26.4%
Literary Prose	362	62.5%	1,732,353	49.1%	24,547,664	49.0%	961,749	39.5%
Philosophy	12	2.1%	39,782	1.1%	586,055	1.2%	73,020	3.0%
Poetry	11	1.9%	18,213	0.5%	252,471	0.5%	68,491	2.8%
Religion	11	1.9%	166,477	4.7%	2,383,475	4.8%	137,240	5.6%
Science	29	5.0%	208,087	5.9%	3,010,912	6.0%	197,685	8.1%
Thoughts	2	0.3%	15,444	0.4%	220,687	0.4%	34,880	1.4%
Total	579		3,527,352		50,086,388		2,436,450	

Table 5. HAC's word distribution in different eras.

Era	Word (Count)	Pct.	Era	Word (Count)	Pct.
Before 600	46,856	2%	1300-1400	3,563,990	7%
600-700	1,001,678	2%	1400-1500	3,508,167	7%
700-800	1,147,709	8%	1500-1600	1,055,423	2%
800-900	4,034,476	11%	1600-1700	857,038	2%
900-1000	5,647,041	11%	1700-1800	762,241	2%
1000-1100	5,653,135	10%	1800-1900	1,188,169	2%
1100-1200	5,153,226	7%	1900-2000	5,416,067	11%
1200-1300	3,638,343	1%	After 2000	7,085,705	14%
Unknown	327,124	2%	*Total*	50,086,388	

5.2 Experimenting with the Historical Arabic Corpus (HAC)

In this experiment, we inquired about the top 100 words and their frequencies in HAC. Table 6 gives the words and their frequencies in the corpus, while Table 7 gives a sample of the top 10 frequent words in each genre. The analysis revealed the following observations:

1. The top words mainly consist of common Arabic verbs, conjunctions and prepositions, reflecting their frequent usage in texts.
2. Words such as "قال" (said) and "قد" (had) indicate narrative and temporal aspects, suggesting a focus on describing events and actions in historical narratives.
3. By examining the top 10 frequent words in each genre, we can observe distinct lexical patterns characteristic of the respective genres, as shown in Table 7. For example, in the historical genre, words related to events, places and individuals are predominant, while in poetry, words associated with emotions, manners, war and environment are more prevalent. In religion, words related to spirituality and faith are predominant, whereas in the scientific genre, words associated with the human body and empirical observations may be more common.
4. Comparing the top frequent words across genres enables researchers to identify similarities and differences in linguistic usage and thematic emphasis. This comparative analysis can highlight genre interrelations and provide a deeper understanding of conventions.

Table 6. Top 100 words and their frequencies in the HAC corpus.

Rank	Word	Freq.	Rank	Word	Freq.	Rank	Word	Freq.	Rank	Word	Freq.
1	من	1,297,204	26	لم	135,987	51	وفي	80,375	76	تعالى	51,155
2	في	1,207,616	27	الذي	135,724	52	كما	78,988	77	فلما	51,053
3	بن	642,586	28	ابو	133,470	53	سنة	78,459	78	كانت	50,997
4	على	582,057	29	فقال	129,347	54	غير	78,258	79	رسول	50,858
5	ان	557,484	30	قد	127,867	55	منه	77,777	80	لها	50,631
6	الى	446,508	31	وقال	124,817	56	مع	75,880	81	مثل	48,512
7	الله	428,560	32	وقد	121,080	57	وسلم	74,301	82	لما	48,211
8	ما	394,041	33	حتى	117,852	58	فيها	74,218	83	ايضا	47,242
9	قال	354,285	34	وكان	116,497	59	الدين	70,727	84	الملك	42,944
10	عن	300,407	35	هو	114,798	60	فإن	70,501	85	يقال	42,791
11	لا	295,900	36	ابي	113,068	61	الناس	69,598	86	قوله	42,607
12	كان	209,242	37	فيه	111,886	62	اليه	69,441	87	عمر	42,592
13	عليه	200,995	38	كل	110,544	63	علي	68,289	88	حدثنا	42,573
14	او	191,781	39	ومن	110,453	64	يكون	67,655	89	احمد	42,484
15	له	191,766	40	انه	109,243	65	وإن	65,896	90	لي	41,093
16	هذا	187,287	41	اي	106,782	66	عنه	64,941	91	ليس	40,320
17	ولا	183,938	42	محمد	106,217	67	ولم	63,116	92	فلا	38,521
18	ثم	182,465	43	التي	103,839	68	عند	62,283	93	هي	38,154
19	ذلك	177,197	44	بين	102,981	69	يقول	59,388	94	قول	37,469
20	اذا	169,685	45	هذه	102,636	70	اهل	57,243	95	قبل	37,227
21	ابن	167,491	46	بعد	87,101	71	وهي	55,793	96	اخر	37,201
22	عبد	154,277	47	صلى	84,441	72	منها	55,364	97	بني	36,283
23	وهو	149,265	48	يا	84,040	73	بعض	54,259	98	فإنه	35,996
24	به	143,352	49	بها	82,465	74	يوم	54,068	99	لأن	35,972
25	الا	142,705	50	وما	81,155	75	فى	52,022	100	شيء	35,805

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 10, No. 04, December 2024.

Table 7. Sample of top 10 words in each genre in the HAC corpus.

History	Literary Prose	Philosophy	Poetry	Religion	Science	Thought
عمر	الزمان	الإنسان	السيف	الله	العين	المجتمع
السلطان	الهوى	العقل	الإبل	محمد	الرأس	النموذج
الحسن	القلب	النفس	الكأس	صلى	المعدة	العلمانية
مدينة	الحب	الفلسفة	الوغى	وسلم	الأدوية	الإنسانية
توفي	القصيدة	الحياة	الليالي	حدثني	الغذاء	الصهيونية
إبراهيم	الأمير	المذهب	الفؤاد	سمعت	البدن	الجماعات
صاحب	ديوان	المنطق	المكارم	أخرجه	الحرارة	المادي
عثمان	الفرزدق	خلدون	الأعداء	هريرة	القلب	النظام
دمشق	ليلى	أفلاطون	المنية	صحيح	السموم	الإمبريالية
بغداد	جارية	الموت	الرماح	رواية	الشيخوخة	الاقتصادي

5.3 Measuring the Linguistic Richness of HAC's Text

Zipf's law, named after the American linguist George Kingsley Zipf [48], who proposed it in the 1930s, can be a helpful tool for identifying important words or concepts in the HAC corpus and measuring its text's lexical richness. Zipf's law, also known as the "law of word frequencies," is an empirical law that describes the statistical distribution of word frequencies in a corpus of natural-language text. The law states that the frequency of a word in a given corpus of text is inversely proportional to its rank in the frequency table. For instance, the second most common word in the corpus will occur approximately a half as often as the most common word, the third most common word will occur approximately one-third as often as the most common word and so forth. Equation (1) can mathematically express the law.

$$f(w) = \frac{k}{r} \tag{1}$$

where f(w) is the frequency of the word w, r is its rank in the frequency table and k is a constant. Similar to English, the law was also observed in Arabic. To show if HAC's content complies with Zipf's law, we inquired about the top 1000 words and their frequencies per each of the eight genres, as shown in Figure 5. We visualized the distribution of words using a log-log scatter chart showing the frequency of each word in the collection plotted against its rank. To assess whether the data aligns with Zipf's law, we applied a linear trendline to find the best-fit straight line. The reliability of this line is highest when the R-squared (R^2) value is close to one. In the case of the entire corpus, the R^2 value was 0.997 and for each of the eight genres, R^2 was close to 0.998. This finding indicates a close adherence of the corpus to Zipf's law.

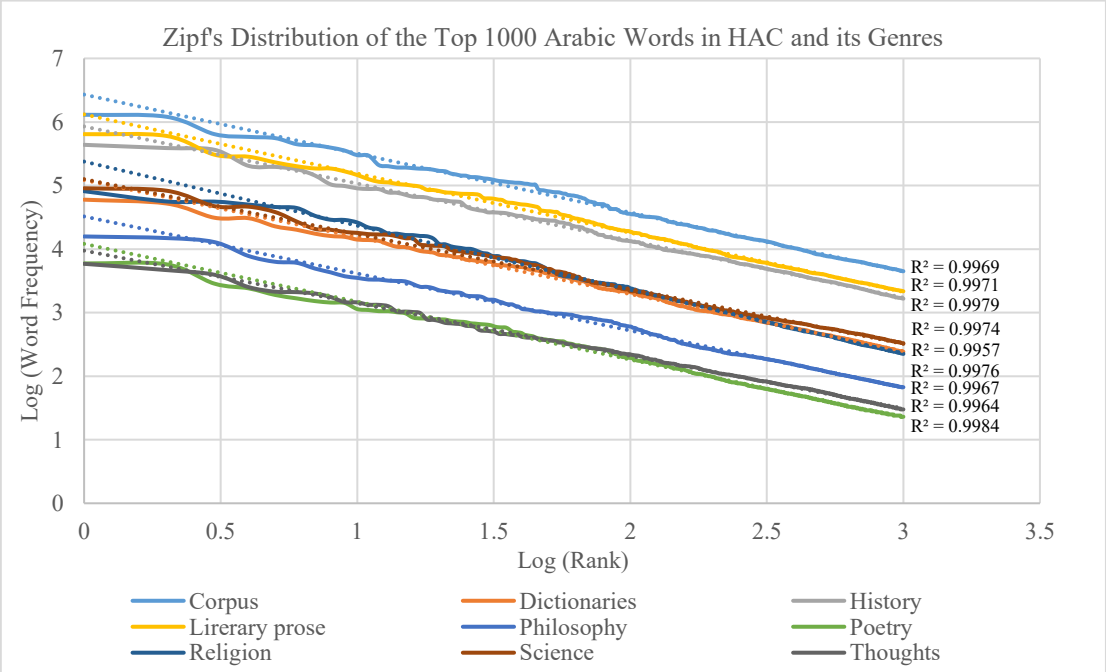


Figure 5. Zipf's distribution of the top 1000 words in the entire HAC corpus and per each of its eight genres.

5.4 Experimenting with the Concordancer

The potential behind developing the new concordancer was to enable students and researchers to explore new research questions and uncover hidden patterns and relationships. In addition, it allows researchers to gain new insights into the history and culture of the Arab world. The concordancer can be used in various fields, including language teaching, translation, linguistic research and text analysis. In general, it allows researchers to (1) explore new research questions and uncover hidden patterns and relationships and (2) gain new insights into the history and culture of the Arab world.

5.4.1 Studying Semantic Change

The phenomenon of semantic change, also referred to as meaning change, is a widespread trend in which the meanings of words undergo alterations over time, either by acquiring new senses or losing old ones, supplanting default senses, shifting in terms of word prototype, narrowing or expanding category boundaries, undergoing pejoration or amelioration and bleaching [3], [49].

Like other languages, Arabic has undergone significant changes over the centuries and many words have shifted meaning to adapt to the needs of every era. Almarwaey and Ahmad [50] suggested that social, economic and political life might influence a word's meaning. Other studies indicated that many words might disappear over time and it becomes necessary to alter the original meanings of dictionaries and language books [1], [51].

Methodology

We conducted a comprehensive corpus analysis to identify and illustrate semantic changes, focusing on high-frequency terms and their historical context. Our approach was designed to ensure a representative selection of terms grounded in systematic criteria.

a) Selection Criteria

1. Frequency and distribution: We identified high-frequency terms across different genres and historical periods within the corpus. Terms were selected based on their consistent presence and significant occurrences, ensuring that they were well-represented across various texts.
2. Diverse representation: The selected terms, كـافر "*kafir*," حجـاب "*hijab*," لحن "*lahn*," and فتنـة "*fitnah*," were chosen to cover a range of semantic fields, including religious, social, cultural and linguistic aspects. This diversity ensures a holistic view of semantic evolution.
3. Historical coverage: We ensured that the chosen terms had a documented presence from early historical eras to contemporary times. This allows for a comprehensive analysis of their semantic trajectories over centuries.

b) Analytical Process

We utilized the concordancer tool to track these terms' usage and semantic shifts over time. This involved several steps:

1. Root-based search: To capture all derivatives and inflections of the selected terms, we conducted searches using their root forms. This approach ensures that variations of each term are included in the analysis, providing a complete picture of their usage.
2. Frequency calculation: The concordancer tool calculated the frequency of each term across different genres and eras, as shown in Tables 8 and 9. This quantitative analysis highlights trends in the popularity and contextual usage of the terms.
3. Contextual analysis: We examined specific instances of the terms in various texts to understand their evolving meanings. Table 10 presents examples of these developments, extracted from different genres and historical periods.

c) Experiment Results and Discussion

The analysis revealed significant semantic shifts for each of the four terms, as follows:

- كافر (*kafir*): Originally meaning "covering" or "concealing," it evolved to signify "disbelief" in the Islamic era, with further contextual variations in the modern period.
- حجـاب (*hijab*): From its early cultural and social dimensions, it has become a contemporary symbol of Islamic identity and women empowerment.
- فتنة (*fitnah*): Transitioned from "temptation" to "political chaos" or "conflict" in modern usage.

- لحن (*lahn*): Shifted from indicating "errors" in speech to representing "melodies" and "music."

Table 10 outlines a few examples of the historical development of the four terms extracted from HAC using the concordancer. As one might notice, the highest distribution of the studied terms was under three main genres: History, Literary Prose and Religion. Let's start with the root *kfr* "كـفـر," which historically meant covering or concealing. The word and its derivatives were spotted in three paragraphs from the year 600 up to the 7th Century, as depicted in Table 9. However, examining the word's derivatives in the Islamic era (7th to 14th Century), it has been spotted in 12,900 paragraphs and has come to signify disbelief or rejection of faith. Over the centuries (19th to mid-20th), the word has been spotted in 3967 paragraphs and was employed in various contexts, including religious discussions, legal matters and cultural discourse. In the contemporary era (Mid-20th to present), the word interpretation can vary widely depending on the cultural, religious and political context.

Table 8. The concordance's statistical search results of four Arabic terms across different genres.

Genre	كفر	حجب	لحن	فتن
History	6145	4568	744	4188
Literary Prose	6588	7560	2575	3569
Philosophy	236	39	20	26
Poetry	56	116	47	29
Religion	3330	348	59	624
Science	559	563	56	352
Thought	45	11	2	19

Table 9. The concordance's statistical search results of four Arabic terms across different eras.

Era	كفر	حجب	لحن	فتن
before 600	3	9	2	0
600-700	1252	220	30	411
700-800	457	174	43	211
800-900	1294	1442	320	526
900-1000	1607	1770	591	682
1000-1100	961	1790	417	572
1100-1200	1789	1428	392	809
1200-1300	1627	1175	266	668
1300-1400	1755	1133	248	746
1400-1500	2158	1080	283	583
1500-1600	435	614	31	224
1600-1700	260	285	51	120
1700-1800	369	273	49	221
1800-1900	256	265	139	358
1900-2000	1297	999	646	1169
after 2000	2414	957	282	1742

Table 10. Samples of the development of four Arabic words from HAC from different genres and eras.

Root	Genre	Era	Text Extracted from the concordance
كفر	Poetry	600-700	الثغر: الطريق في الجبل. **الكافر**: الليل الذي يستر كل ما يقع عليه…
	Poetry	600-700	في ليلةٍ **كَفَرَ** النُّجومَ غَمَامُهَا (8) _ (1) السري: النهر الصغير…
	Literary prose	1000-1100	و**الكفر** مجتمع على الإيمان وضاقت الطرق بكثرة الرماح وأهل **الكفر** …
	Literary prose	1400-1500	يَتَبَدَّلِ **الْكُفْرَ** بِالْإِيمانِ فَقَدْ ضَلَّ سَواءَ السَّبِيلِ «2» . فقلت: يا شيخ…
	History	1500-1600	واقتتلوا مَعَهم وَقتل جمَاعَة من **الكفَّار** وَاسْتَشْهدَ ثَلَاثَة من المماليك الْخَواص…
حجب	Poetry	900-1000	(جعل ابن حزم **حاجبا** .. سُبحانَ من جعلَ ابنَ حزم **يحجب**) وقال آخر: (**إحتجب** الكاتب…
	History	1000-1100	**الحاجب**: وهو الذي يقف على باب القاضي، **ليحجب** عنه الناس أثناء النظر في الدعاوي…
	History	1900-2000	الالتزام التام **بالحجاب** الإسلامي حيث أن المحادثات بينهما كانت تتم بواسطة امرأة تُندب لهذا الأمر،…
	History	after 2000	وعملت حكومته على إلغاء **حجاب** المرأة وأمرت بالسفور…
فتن	Literary prose	800-900	بك والصبر عنك ما لا يكون يا غزالاً بلحظه **يفتن** الناس وفي طرفه…
	Literary prose	800-900	**الفتنة** في هذا الموضع: النعمة واللذة. ومنه قول الله جل وعز: (إنما أموالكم وأولادكم **فتنة**…
	Literary prose	1900-2000	ما يسمى **بالفتن** وثورات الطامحين والمنشقين عن طاعة قرطبة، ويكفي أن…
	History	after 2000	على المسلمين باب **الفتنة** إلى اليوم. وهذا الورع الجاهل نلاحظه اليوم في تصرفات بعض المسلمين…
لحن	Literary prose	800-900	وَمِمَّنْ كَانَ لَا **يُلْحَن** أَلْبَتَّهُ حَتَّى كَأَنَّ لِسَانه أَعْرَابِيّ فَصِيح أَبُو زَيْد الْنَّحْوِيّ وَأَبُو سَعِيد…
	Literary prose	800-900	وهي الايجاز والابتعاد عن **اللحن**، ووضوح المعنى واللفظ، وعدم اللجوء الى الزخرفة البيانية،…
	Literary prose	1900-2000	وفي فمي **لحن** وشهد وراح فالراح في البيت الأخير على العكس من الراح…
	Literary prose	after 2000	إيقاع واحد، والجمع: أناشيد. وإن كان الإنشاد للشعر قد يصحبه **تلحين** وحسن إيقاع،…

Similarly, Table 10 shows the root *hjb* "حجب" across different eras, from its conceptual origins in the pre-Islamic era to its various cultural and social dimensions in later eras. In its original meaning, *hajaba* meant "to hide" and the role of *hajeb* was prestigious in the Islamic-Caliphate periods. The contemporary era has seen a revival in the use of the *hijab* as a symbol of Islamic identity, modesty, faith and empowerment for many women. Initially, the root *ftn* "فتن" meant "temptation". However, the meaning of *fitna* in contemporary Arabic usage has shifted to be associated with political chaos or conflict over time. Some other words have undergone semantic elevation, moving from negative meanings to ones that are now significant and positive. For example, the root *lhn* "لحن" was used to signify errors or discord in speech and it has transformed to refer to the sweetness of melodies and music.

These are just a few examples of the many Arabic words that have changed meaning over time. These changes in meaning can be challenging for researchers and scholars working with historical Arabic texts, as they may need to consider the historical context to interpret the definition of a word or phrase accurately. Additionally, they may need to be aware of these changes in meaning when comparing historical texts to contemporary usage.

These findings, supported by the data in Tables 8, 9 and 10, illustrate the comprehensive corpus analysis and ensure that our study provides a robust historical corpus and tools to advance research in Arabic linguistics and NLP.

6. Conclusion and Future Work

The Historical Arabic Corpus (HAC) and the developed tools provide an excellent resource for extracting historical semantic knowledge. The importance of HAC lies in its rich and extensive history, spanning over 1600 years and providing a unique perspective on the development of the historical Arabic context. These texts are a valuable resource for scholars and researchers seeking to study the Arabic language and to understand the Islamic world's cultural, social and political contexts and have contributed significantly to the field of history.

The HAC corpus can be a valuable resource for both native speakers and foreign learners of Arabic-language learning. For native speakers, HAC offers opportunities to explore classical Arabic texts, deepen their linguistic proficiency and engage with cultural heritage. On the other hand, foreign learners can utilize HAC to enhance their understanding of classical Arabic vocabulary and immerse themselves in historical contexts. Instructors can incorporate HAC into lesson plans by assigning readings, conducting comparative analyses between modern and historical Arabic texts and guiding discussions on linguistic evolution. Learners can independently utilize HAC for vocabulary expansion, comprehension tasks and research projects on specific historical periods.

We have created a collection of tools for handling and experimenting with HAC. The corpus builder incorporates a stemmer and a tagger to annotate and manipulate documents and save them in a database. We tokenized and normalized the corpus words into an indexer for efficient searching. We also created an easy-to-use concordancer to assist in searching and extracting linguistic knowledge from HAC, as well as helping in compiling dictionary entries for a hypothetical historical dictionary.

Our goal is to improve HAC by providing more accurate annotation, enlarging the corpus to represent Arabic more thoroughly, optimizing and adding features to the search engine and the concordancer, responding to the needs of linguists and offering more flexibility to meet their satisfaction.

Acknowledgments

This research was completed while the first author was on sabbatical leave from The University of Jordan for the academic year 2021/2022 to Princess Sumaya University for Technology, Amman, Jordan. The authors would like to thank Omaima Ismail for her initial contribution to the HAC project.

References

[1] R. Laatar, C. Aloulou and L. Hadrich Belguith, "Towards a Historical Dictionary for Arabic Language," International Journal of Speech Technology, vol. 25, no. 1, pp. 29-41, 2022.

[2] B. Hammo, S. Yagi, O. Ismail and M. Abushariah, "Exploring and Exploiting a Historical Corpus for Arabic," Language Resources & Evaluation, vol. 50, pp. 839-861, DOI:10.1007/s10579-015-9304-9, 2016.

[3] O. Ismail, S. Yagi and B. Hammo, "Corpus Linguistic Tools for Historical Semantics in Arabic," International Journal of Arabic-English Studies, vol. 15, pp. 135-152, 2014.

[4] E. Al-Thwaib, B. H. Hammo and S. Yagi, "An Academic Arabic Corpus for Plagiarism Detection: Design, Construction and Experimentation," Int. Journal of Educational Technology in Higher Education, vol. 17, no. 1, DOI:10.1186/s41239-019-0174-x, 2020.

[5] A.F. Mukhamadiarova, "Application of Corpus-based Technologies in the Formation of Lexical and Grammatical Skills in German," Perspectives of Science and Education, vol. 53, pp. 247-259, DOI:10.32744/pse.2021.5.17, 2021.

[6] S. P. Cheng, "University Students' Perceived Benefits and Difficulties Related to Corpus-assisted Translation," Compilation and Translation Review, vol. 16, no. 1, pp. 81-132, 2023.

[7] A. Boulton, "Data-driven Learning: Taking the Computer out of the Equation," Language Learning, vol. 60, no. 3, pp. 534-572, 2010.

[8] L. Zhao, W. Kong and C. Wang, "Electricity Corpus Construction Based on Data Mining and Machine Learning Algorithm," Proc. of the 2020 IEEE 5th Information Technology and Mechatronics Engineering Conf. (ITOEC), pp. 1478-1481, Chongqing, China, 2020.

[9] A. O'keeffe, M. McCarthy and R. Carter, From Corpus to Classroom: Language Use and Language Teaching, DOI: 10.1017/CBO9780511497650, Cambridge University Press, 2007.

[10] T. Wambsganss, T. Kueng, M. Soellner and J. M. Leimeister, "ArgueTutor: An Adaptive Dialog-based Learning System for Argumentation Skills," Proc. of the 2021 CHI Conf. on Human Factors in Computing Systems, pp. 1-13, DOI: 10.1145/3411764.3445781, May 2021.

[11] Essential Corpus Tools, [Online], Available: https://corpus-analysis.com/, Last visited in June 2024.

[12] S. Khoja, "Khoja's Stemmer," [Online], Available: http://zeus.cs.pacificu.edu/shereen/research.htm, 2015. Accessed April 2024.

[13] K. Toutanova, D. Klein, C. Manning and Y. Singer, "Feature-rich Part-of-speech Tagging with a Cyclic Dependency Network," Proc. of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL 2003), pp. 252-259, 2003.

[14] A. Abdelali, K. Darwish, N. Durrani and H. Mubarak, "Farasa: A Fast and Furious Segmenter for Arabic," Proc. of the 2016 Conf. of the North American Chapter of the Association for Computational Linguistics: Demonstrations, pp. 11-16, San Diego, California, 2016.

[15] O. Ossama, N. Zalmout, S. Khalifa, D. Taji, M. Oudah, B. Alhafni, G. Inoue, F. Eryani, A. Erdmann and N. Habash, "CAMeL Tools: An Open Source Python Toolkit for Arabic Natural Language Processing," Proc. of the 12th Language Resources and Evaluation Conf., LREC, pp. 7022-7032, Marseille, France, 2020.

[16] J. Zare, S. Karimpour and K. Aqajani Delavar, "Classroom Concordancing and English Academic Lecture Comprehension: An Implication of Data-driven Learning," Computer Assisted Language Learning, vol. 36, nos. 5-6, pp. 885-905, DOI:10.1080/09588221.2021.1953081, 2023.

[17] J. Zare, S. Karimpour and K. A. Delavar, "The Impact of Concordancing on English Learners' Foreign Language Anxiety and Enjoyment: An Application of Data-driven Learning," System, vol. 109, p. 102891, DOI:10.1016/j.system.2022.102891, 2022.

[18] V. Mohammadi and N. Mohit, "Student and Teacher Attitude toward Using Concordancing in Learning and Teaching Preposition Collocations: Issues and Options," Journal of Language Horizons, vol. 5, no. 2, pp. 139-166, 2021.

[19] I. Kazaz, "Alternative Vocabulary Assessment: Using Concordance Line Activities for Testing Lexical Knowledge," Int. Online Journal of Education and Teaching, vol. 7, no. 3, pp. 1221-1238, 2020.

[20] A. T. Shawaqfeh and M. A. Khasawneh, "Incorporating Corpus Linguistics Tools in the Training and Professional Development of Lecturers in Translation Studies," Studies in Media and Communication, vol. 11, no. 7, p. 260, DOI:10.11114/smc.v11i7.6379, 2023.

[21] M. del Mar Sánchez Ramos, "Teaching English for Medical Translation: A Corpus-based Approach," Iranian Journal of Language Teaching Research, vol. 8, no. 2, pp. 25-40, 2020.

[22] O. J. Ballance and A. Coxhead, "How Much Vocabulary is Needed to Use a Concordance?" Int. Journal of Corpus Linguistics, vol. 25, no. 1, pp. 36-61, DOI:10.1075/ijcl.17116.bal, 2020.

[23] S. Un-udom and N. Un-udom, "A Corpus-based Study on the Use of Reporting Verbs in Applied Linguistics Articles," English Language Teaching, vol. 13, no. 4, pp. 162-169, 2020.

[24] M. Bednarek and G. Carr, "Computer-assisted Digital Text Analysis for Journalism and Communications Research: Introducing Corpus Linguistic Techniques That Do not Require Programming," Media International Australia, vol. 181, no. 1, pp. 131-151, DOI: 10.1177/1329878X20947124, 2021.

[25] A. Eddakrouri, "Arabic Corpus of Library and Information Science: Design and Construction," Egyptian Journal of Language Engineering, vol. 10, no. 1, pp. 1-9, DOI:10.21608/ejle.2023.183529.1040, 2023.

[26] S. Khoja, "An RSS Feed Analysis Application and Corpus Builder," Proc. of the 2nd Int. Conf. on Arabic Language Resources and Tools, pp. 115-118, Cairo, Egypt, 2009.

[27] M. O'Donnell, "The UAM Corpus Tool: Software for Corpus Annotation and Exploration," Proc. of Bretones Callejas, Carmen M. et al. (eds.) Applied Linguistics Now: Understanding Language and Mind, Almería: Universidad de Almería, pp. 1433-1447, 2008.

[28] S. Alansary, M. Nagi and N. Adly, "Towards Analyzing the International Corpus of Arabic (ICA): Progress of Morphological Stage," Proc. of the 8th Int. Conf. on Language Engineering, Cairo, Egypt, 2008.
[29] M. Attia, P. Pecina, L. Tounsi, A. Toral and J. van Genabith, "Lexical Profiling for Arabic," Proc. of eLex 2011, pp. 23-33, 2011.
[30] K. Dukes and N. Habash, "Morphological Annotation of Quranic Arabic," Proc. of the 7th Int. Conf. on Language Resources and Evaluation (LREC'10), pp. 2530-2536, Valletta, Malta, 2010.
[31] M. Boella, F. Romani, A. Al-Raies, C. Solimando and G. Lancioni, "The SALAH Project: Segmentation and Linguistic Analysis of Ḥadīṯ Arabic Texts," Proc. of Information Retrieval Technology, Part of the Book Series Lecture Notes in Computer Science, vol. 7097, pp. 538-549, DOI: 10.1007/978-3-642-25631-8_49, Springer, Berlin, Heidelberg, 2011.
[32] A. Sharaf and E. Atwell, "QurAna: Corpus of the Quran Annotated with Pronominal Anaphora," Proc. of the 8th Int. Conf. on Language Resources and Evaluation (LREC'12), pp. 130-137, Istanbul, Turkey, 2012.
[33] S. Altammami, E. Atwell and A. Alsalka, "Constructing a Bilingual Hadith Corpus Using a Segmentation Tool," Proc. of the 12th Language Resources and Evaluation Conf., pp. 3390-3398, Marseille, France, 2020.
[34] M. Hajjar, A. Al-Hajjar, K. Zreik and P. Gallinari, "An Improved Structured and Progressive Electronic Dictionary for the Arabic Language: iSPEDAL," Proc. of the 5th Int. Conf. on Internet and Web Applications and Services (ICIW), pp. 489-495, Barcelona, Spain, 2010.
[35] B. Hammo, F. Al-Shargi, S. Yagi and N. Obeid, "Developing Tools for Arabic Corpus for Researchers," Proc. of the 2nd Workshop on Arabic Corpus Linguistics (WACL-2), Lancaster University, UK, 2013.
[36] Z. Alyafeai, M. Masoud, M. Ghaleb and M. S. Al-shaibani, "Masader: Metadata Sourcing for Arabic Text and Speech Data Resources," Proc. of the 13th Language Resources and Evaluation Conf., pp 6340–6351, European Language Resources Association, Marseille, France, 2022.
[37] The Linguistic Data Consortium, [Online], Available: https://www.ldc.upenn.edu/sites/www.ldc.upenn.edu/files/arabic.pdf. (Last visited in June 2024).
[38] M. Alrabiah, A. Al-Salman and E. Atwell, "The Design and Construction of the 50 Million Words KSUCCA," Proc. of the 2nd Workshop on Arabic Corpus Linguistics (WACL-2), Lancaster University, UK, 2013.
[39] A. O. Al-Thubaity, "A 700 M + Arabic Corpus: KACST Arabic Corpus Design and Construction," Language Resources and Evaluation, vol. 49, pp. 721-75, DOI: 10.1007/s10579-014-9284-1, 2015.
[40] L. Anthony, "AntConc: Design and Development of a Freeware Corpus Analysis Toolkit for the Technical Writing Classroom," Proc. of Professional Communication Conf. (IPCC 2005), pp. 729-737, Limerick, Ireland, 2005.
[41] A. Roberts, L. Al-Sulaiti and E. Atwell, "aConCorde: Towards an Open-source, Extendable Concordancer for Arabic," Corpora, vol. 1, no. 1, pp. 39-60, 2006.
[42] R. Abbès and J. Dichy, "AraConc, an Arabic Concordance Software Based on the DIINAR.1 Language Resource," Proc. of the 6th Int. Conf. on Informatics and Systems, pp. 127-134, 2008.
[43] Y. Belinkov, A. Magidow, M. Romanov, A. Shmidman and M. Koppel, "Shamela: A Large-scale Historical Arabic Corpus," Proc. of the Workshop on Language Technology Resources and Tools for Digital Humanities (LT4DH), pp. 45–53, Osaka, Japan, 2016.
[44] A. Hourani, A History of the Arab Peoples: Updated Edition, ISBN: 9780571288014, London: Faber and Faber, 2013.
[45] L. Larkey and M. Connell, "Arabic Information Retrieval at UMass in TREC-10," Proc. of the 10th Text Retrieval Conference (TREC-10), pp. 562-570, Maryland, USA, 2010.
[46] R. H. Al Mahmoud, B. H. Hammo and H. Faris, "Cluster-based Ensemble Learning Model for Improving Sentiment Classification of Arabic Documents," Natural Language Engineering, pp. 1-39, DOI: 10.1017/S135132492300027X, 2023.
[47] R. H. AlMahmoud and B. H. Hammo, "SEWAR: A Corpus-based N-gram Approach for Extracting Semantically-related Words from Arabic Medical Corpus," Expert Systems with Applications, vol. 238, p. 121767, DOI: 10.1016/j.eswa.2023.121767, 2014.
[48] G. K. Zipf, "The Meaning-Frequency Relationship of Words," The Journal of General Psychology, vol. 33, no. 2, pp. 251-256, DOI:10.1080/00221309.1945.10544509, 1945.
[49] N. N. Hanifah, "The Origin of Arabic Lexicography: Its Emergence and Evolution," HuRuf Journal: Int. Journal of Arabic Applied Linguistic, vol. 1, no. 2, pp. 238-251, DOI: 10.30983/huruf.v1i1.4932, 2021.
[50] A. O. Almarwaey and U. K. Ahmad, "Semantic Change of Hijab, Halal and Islamist from Arabic to English. 3L: Language, Linguistics, Literature," The Southeast Asian Journal of English Language Studies, vol. 27, no. 2, pp. 161-176, DOI: 10.17576/3L-2021-2702-12, 2021.
[51] R. Laatar, A. Rhayem, C. Aloulou and L. H. Belguith, "Towards a Historical Ontology for Arabic Language: Investigation and Future Directions," Proc. of the Int. Conf. on Intelligent Systems Design and Applications, Part of the Book Series: Lecture Notes in Networks and Systems, vol. 418, pp. 1078-1087, Cham: Springer International Publishing, December 2021.

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 10, No. 04, December 2024.

ملخص البحث:

تستكشف هذه الورقة تطوير وتصميم وإعادة بناء مجموعة نصوص تاريخية باللّغة العربية (HAC) تُغطّي فترة زمنية تتجاوز 1600 سنة من الاستخدامات اللُّغوية غير المنقطعة. وتؤكّد الورقة الجوانب التّقنية المتّبعة لتحسين النّظام، وتعرض التّجارب الّتي تمّ إجراؤها على مجموعة النّصوص. هذا مع الإشارة إلى أنّ اللُّغة العربية تمتلك إرثاً أدبياً وثقافياً يمتدّ لآلاف السّنين. وقد جعل شمول المصادر الرّقمية بالإضافة إلى التّقدّم في تقنيات معالجة اللّغات الطّبيعية مجموعات النّصوص التّاريخية باللّغة العربية ذات أهمّية متزايدة للباحثين والمتعلمين في جميع أنحاء العالم.

ومن خلال دمْج مجموعة النُّصوص باللُّغة العربية وأدواتها في تعلُّم اللغة العربية، يُمْكن للمعلِّمين أن يغوصوا على نحوٍ أعمق في المصْطَلحات اللُّغوية والثّقافة العربية ويكتسبوا نظرةً ثاقبةً في سبيل تحسين المهارات اللغوية وفهْم اللُّغة العربية.

والجدير بالذّكر أنّ دمْج الإرشادات البشرية وتقنيات معالجة اللُّغات الطّبيعية من شأنه أن يجعل تعلُّم اللُّغة أمراً مُمتعاً، وذلك من خلال تقديم طريقةٍ ديناميكية وأصيلة لإتقان اللُّغة العربية.

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 10, No. 04, December 2024.

IMAGE ENCRYPTION TECHNIQUE BASED ON BINARY COMBINATION OF MULTIPLE CHAOTIC MAPS AND DNA SEQUENCE OPERATIONS

Nisreen I. R. Yassin

(Received: 11-May-2024, Revised: 17-Jul.-2024, Accepted: 30-Jul.-2024)

ABSTRACT

The huge advance of digital communication and networks has led to enormous storage and transmission of information over public networks. Nevertheless, the assurance of information security remains incomplete across these unsecured networks. Currently, digital images are the primary mean for sharing information over open networks. Consequently, the confidentiality of digital images during storage and transmission has become a crucial concern, particularly when sharing sensitive information. Image encryption has emerged as a solution to this problem. This paper presents an image encryption technique based on multiple one-dimensional chaotic maps and DNA coding. The technique employs three one-dimensional chaotic maps, including the logistic map, tent map and piecewise map, multiple times to produce 18 random sequences with different initial values and parameters. SHA-512 hash function is used to indicate the initial values of chaotic maps. For encrypting images, the binary elements from various sequences of chaotic maps are amalgamating to alter the pixel intensities of the image in the diffusion process. Dynamic DNA coding is performed through random selection of DNA rules and operations (XOR, XNOR and Addition) to each pixel in the image. The technique is enforced using circular rotations which are applied randomly to each key. The proposed technique is evaluated using many standard images. Different performance metrics have been measured. The empirical findings illustrate the security and resilience of the suggested method and its ability to resist statistical and differential attacks.

KEYWORDS

Information security, Cryptography, Chaotic maps, Image encryption, DNA.

1. INTRODUCTION

The transmission and sharing of information and confidential data mostly occur *via* digital networks and the internet, using digital images, texts, voice messages and videos. This transmission includes online banking, e-commerce transactions and the military, which has increased the need for data protection and privacy preservation. Of all the data types, digital images have the ability to efficiently communicate visual information with compact storage sizes, thus they are the most commonly used on social applications [1]. Digital images are widely shared through various communication media, so it is essential to ensure the security of these images from malicious and unlicensed access attempts [2]-[3].

Cryptography is one of the information-security techniques used by many security companies to protect social data, government communications and personal transactions. Cryptography uses a secret key to convert data into an unreadable format that has no textual or visual meaning. In this case, the attacker cannot access the content directly, but tries to find the key or disrupt the access process. Therefore, the basic goal is to increase the computational cost of the unauthorized decryption process to make it infeasible [4]-[5]. Classical encryption algorithms such as Data Encryption Standard (DES) [6] or Advanced Encryption Standard (AES) [7] which are used for text encryption have demonstrated ineffectiveness in encrypting images due to strong image pixel correlation, high redundancy and the large size of images [8]-[9]. For instance, when an image undergoes encryption *via* a substitution cipher, it merely alters the colour of the image, allowing retrieval or comprehension by a malicious entity. Image encryption depends on two processes which are confusion and diffusion. The confusion process aims to change the locations of image pixels, thus breaking the correlation between adjacent pixels and each pixel in the cipher image will be related to a part of the used key. The diffusion process aims to alter the image-pixel values, therefore diffusing the frequencies of the plain-image pixels [10].

Chaos systems are nonlinear, deterministic, but unpredictable systems, first introduced by Matthews for image encryption [11]-[12]. Chaos systems are characterized by strong randomness and high sensitivity to parameters and initial conditions. Chaotic maps are one-dimensional or high-dimensional maps.

N. Yassin is with National Research Centre, Cairo, Egypt. Email: nisreen.yassin20@gmail.com

One-dimensional chaotic maps are easy to implement in software and hardware due to their simple structure. The cryptosystems based on one- or two-dimensional chaotic maps are characterized by simplicity and low execution time, but the defects are small key space and restricted complexity, which lead to soft security [13]-[15]. On the other hand, high-dimensional chaotic maps have a larger number of initial conditions and control parameters, which leads to a larger key space and thus more security [16]. The drawback of these maps is the increased execution time required to solve high-dimensional equations. As a solution, researchers used a combination of simple chaotic maps instead of high-dimensional maps [3].

Scientists have found that using only chaotic maps for image encryption is not enough to achieve high security. DNA coding has been combined with chaotic maps to improve the diffusion of the image-encryption process. Using this mixture has led to promising results according to randomness and nonlinearity. This study aims to leverage the characteristics of one-dimensional chaotic maps, specifically the simplicity of their construction and their pronounced sensitivity to initial conditions. Meanwhile, it seeks to address the limitations linked to these properties, including a limited key space and constrained complexity. The idea is to generate multiple maps with different initial values and different control parameters, therefore achieving the necessity of large key space. SHA- 512 hash function and DNA coding are combined with chaotic maps to increase security and enhance the diffusion process. In this paper, a novel image-encryption technique is proposed. The technique is a pixel-based technique that uses three one-dimensional chaotic maps (Logistic Map, Tent Map and Piecewise Map) multiple times to generate random sequences required for permutation and diffusion processes. The keys of the diffusion process are formed by mixing the corresponding binary digits of the random sequences followed by a circular rotation performed randomly according to a random-number generator. DNA coding for image pixels and generated keys is performed using a randomly selected rule number. Then, a randomly selected operation of the three DNA mathematical operations (XOR, XNOR and Addition) is performed. To enhance security, XOR is applied with another mixed key to obtain the final cipher image.

The main contributions of the proposed work are as follows:

- The manuscript introduces a resilient and efficient method for encrypting images by amalgamating the binary elements from various sequences of chaotic maps to alter the pixel intensities in the image.
- Erratic DNA coding is performed using randomly selected DNA rule numbers and DNA mathematical operations.
- The technique is reinforced by using the SHA-512 hash function and randomly circular rotations.
- The technique is designed to have multiple initial conditions and control parameters and a large key space.

The remainder of the paper is arranged as follows: related works are introduced in Section 2, preliminary works are given in Section 3, the proposed method is given in Section 4, the analysis results are shown in Section 5 and conclusions are stated in Section 6.

2. RELATED WORK

Image encryption techniques depend mainly on two phases, the confusion phase and the diffusion phase. In this section, some related state-of-the-art techniques are presented. Qobbi et al. [17] presented an image-encryption system based on two 1D chaotic maps, which are logistic and tent maps. These maps are used to generate permutation and substitution table. The proposed system is highly related to the plain image. Kumar and Girdhar [18] developed a scheme for image encryption based on DNA coding and chaotic maps. Lorenz and Rossler chaotic maps are used to diffuse the image at the pixel level, then a 2D logistic map is performed at the bit level to confuse the image. Rahul et al. [19] used the logistic map, the Henon map and the Lorenz system associated with DNA encoding to safeguard digital images. Their approach included enhancing the system through the utilization of the SHA-256 hash function and zigzag traversal. Akraam et al. [3] proposed an encryption algorithm based on a combination of the chaotic maps: logistic map, piecewise linear chaotic map, tent map and Henon map. These maps are used to generate two keys which are used to diffuse the decimal pixels of the image. Li et al. [20] used a classical chaotic map and a two-dimensional Lorenz chaotic map to encrypt and decrypt the image.

Niu et al. [21] proposed an image-encryption scheme based on chaotic maps and DNA. They used logistic and Henon maps to perform pixel permutation and diffusion. Bidirectional exclusive OR operations are used to strengthen the scheme. Ibrahim et al. [14] used a logistic chaotic function to generate random round keys, which are used to modify individual pixels using the DNA Playfair matrix. Wu et al. [22] proposed a new map called two-dimensional Hénon-Sine map. DNA encoding and XOR function are applied to encrypt the image. Gera and Agrawal [23] proposed an image-encryption approach based on a 4D discrete hyperchaotic map and DNA coding. Global scrambling is performed on the binary image, then DNA algebraic operations are applied for diffusion. Wan et al. [24] used a modular operation to form a new one-dimensional chaotic map based on several existing one-dimensional chaotic maps. The new chaotic map combined with DNA coding is used to encrypt the image. Enayatifar et al. [25] proposed an encryption system based on DNA coding and logistic map and using a genetic algorithm to indicate the best DNA mask for encryption. Liu et al. [26] introduced a novel image-encryption method designed to encrypt numerous medical images simultaneously. This proposed approach relies on a recent chaotic model and a novel DNA operation. Lai et al. [27] introduced a novel memristive Hopfield neural network (HNN) which was utilized for the creation of an image-encryption scheme. Lai et al. [28] introduced a theoretical structure aimed at the creation of the ultraboosting memristive hyperchaotic map. The framework gives rise to four distinct hyperchaotic maps, all of which are utilized in the design of an image-encryption technique. Liu et al. [29] proposed an encryption technique to encrypt remotely sensed airport images. The technique is based on indicating the positions of sensitive information then encrypting these positions using high-speed index-dynamic diffusion. Liu et al. [30] enhanced sine cross coupled mapping lattice (ISCCML) to produce a better key stream. Moreover, a fractal disordered matrix (FDM) is introduced, which exhibits iterative and out-of-order characteristics, designed for the concurrent scrambling diffusion of images. Wang et al. [31] proposed an image-encryption scheme based on two-parameter wide-range system with a mixed coupled map lattice model (TWMCML). Three-dimensional bit-level coupled XOR technique is used to scramble the significant regions of the image. Liu et al. [32] proposed a chaotic system called improved sinusoidal dynamic non-adjacent coupled mapping lattice (ISDNCML). Based on the proposed chaotic system, the private and non-private regions of an image have been encrypted. Lia and Liu [33] derived a 2D-hyperchaotic map used to encrypt colour images. Strong encryption is obtained by using circular-shift confusion and bidirectional-parallel diffusion.

3. PRELIMINARY WORKS

3.1 Chaotic Maps

Chaotic maps are mathematical functions that generate a highly erratic pattern based on the initial seed value. These dynamic systems can generate millions of bits in a period before the pattern repeats with ultimate sensitivity to initial conditions or changes in control parameters. This feature of chaotic maps leads to their extensive use in multimedia encryption, where a hacker cannot implement pattern-analysis attacks [34]. This section describes three chaotic maps used in the proposed technique.

3.1.1 Logistic Map

The logistic map is characterized by its simplistic nature, yet intricate dynamical behaviour. It is defined by Equation (1) [10], [17].

$$X_n = rX_{(n-1)}\left(1 - X_{(n-1)}\right) \tag{1}$$

where X_n is the population at time n and $0 < X_n < 1$. r is the control parameter that lies in the interval $[0\ \ 4]$. The logistic map exhibits chaotic behaviour for r values exceeding 3.569945. Figure 1(a) shows the bifurcation diagram of the logistic map using control parameters in the range of [3 4].

3.1.2 Tent Map

The tent map is a simple chaotic map calculated using Equation (2) [17], [35].

$$T_{n+1} = \begin{cases} uT_n, & if \quad T_n < 0.5 \\ u(1 - T_n), & otherwise \end{cases} \tag{2}$$

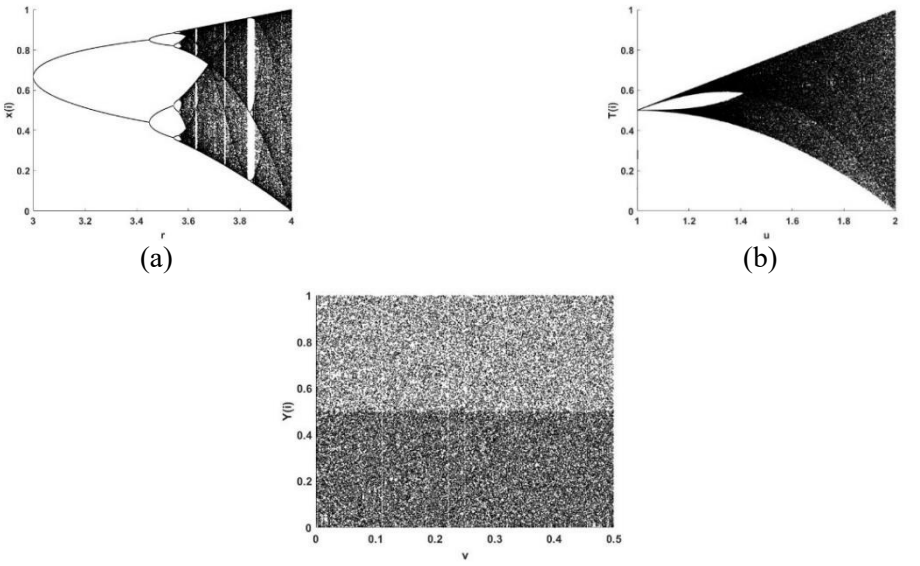
where $0 \leq u \leq 2$ is the control parameter and $T_n \in [0\ \ 1]$. T_0 is the initial value. Figure 1(b) shows the bifurcation diagram of the tent map using control parameters in the range of [1 2].

3.1.3 Piecewise Linear Chaotic Map

The mathematical representation of the piecewise linear chaotic map is shown in Equation (3) [36].

$$Y_{n+1} = \begin{cases} \frac{Y_n}{v}, & if \quad 0 \le Y_n < v \\ \frac{Y_n - v}{0.5 - v}, & if \quad v \le Y_n < 0.5 \\ 1 - Y_n, & if \quad 0.5 \le Y_n < 1 \end{cases} \tag{3}$$

The initial value Y_0 is in the interval $[0 \ \ 1]$ and $v \in [0 \ \ 0.5]$ is the control parameter. Figure 1(c) shows the bifurcation diagram of the piecewise linear chaotic map using control parameters in the range of [0 0.5].



(a) (b) (c)

Figure 1. Bifurcation diagrams of (a) logistic map, (b) tent map and (c) piecewise linear chaotic map.

3.2 DNA Encoding

Deoxyribonucleic acid (DNA) is a chemical compound that is responsible for generating all types of cell proteins in living organisms. It consists of four nucleic acid bases: adenine (A), thymine (T), guanine (G) and cytosine (C). The relationship between these bases is elucidated by the Watson-Crick rules [37], which posit that T (11) is the complement of A (00) and C (01) is the complement of G (10). There are eight rules that represent the binary encoding of DNA nucleotides, where the utilized binary bases are 00, 11, 01 and 10. The binary representation of DNA encoding rules is illustrated in Table 1. The proposed technique employs the use of DNA mathematical operations, especially XOR, XNOR and Addition, during the diffusion process. The DNA representation of these operations is presented in Table 2.

Table 1. DNA encoding rules.

Rule	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
A	00	00	01	01	10	10	11	11
T	11	11	10	10	01	01	00	00
C	01	10	00	11	00	11	01	10
G	10	01	11	00	11	00	10	01

Table 2. DNA mathematical functions.

	XOR				**XNOR**				**Addition**			
	A	**T**	**C**	**G**	**A**	**T**	**C**	**G**	**A**	**T**	**C**	**G**
A	A	T	C	G	T	A	G	C	A	T	C	G
T	T	A	G	C	A	T	C	G	T	C	G	A
C	C	G	A	T	G	C	T	A	C	G	A	T
G	G	C	T	A	C	G	A	T	G	A	T	C

4. THE PROPOSED TECHNIQUE

The proposed technique comprises three distinct stages: key generation, permutation and diffusion. The key-generation process depends on the SHA 512 hash function of the plain image to compute the initial conditions of the chaotic maps. Therefore, any change in the plain image directly affects the algorithm keys. The permutation process involves the random positioning of pixels of the original image according to the coordinates of chaotic-map sequences. The diffusion process entails the mixing of binary bits from a set of generated chaotic maps and DNA coding. The detailed steps of the proposed technique are as follows.

1- The initial step is to read the plain image I of size MxN, where M and N are the dimensions of the image in terms of rows and columns, respectively.

2- Generate the SHA 512 hash function of I to be used as a unique key. Convert it into a binary sequence B of length l_B. Divide B into a number of parts P according to the required number of initial values ni as illustrated in Equation (4).

$$P = \{P_1,\ P_2, \dots, P_{ni}\} = \lfloor (l_B/ni) \rfloor \tag{4}$$

where l_B is the length of SHA 512 hash function which is 512 bits and ni =18.

3- Divide each part p into groups of eight bits, as indicated by Equation (5). The initial conditions of the chaotic maps, designated as x_i, are calculated. In the proposed technique, the number of required initial conditions is 18, therefore 18 keys are generated by XOR bits from each group.

$$K(i) = \left\{(k_1 \oplus k_2 \oplus k_3)_{P_1}, \dots, (k_1 \oplus k_2 \oplus k_3)_{P_{ni}}\right\} = \lfloor \{P_1,\ P_2, \dots, P_{ni}\}/8) \rfloor \tag{5}$$

$$x_i = K(i)/256 \tag{6}$$

where each k is a fold of 8 bits and i is a number from 1 to n_i.

4- Reshape I into one-dimensional array I_1 of size $L = 1xMN$. For RGB images, concatenate the three colour channels R, G and B of the image I to be an image of size $Mx3N$, then reshape I into one-dimensional array I_1 of size $L = 1x3MN$.

5- Generate two chaotic sequences $X_1$and T_1 of size L using logistic map and tent map illustrated by Eq. (1) and Eq. (2) using two initial values calculated by Equation (6).

6- Sort sequences X_1 and T_1 ascendingly to get the positions of their members P_1 and P_2. The two arrays P_1 and P_2 contain L random numbers, which will be used for permutation the image I_1.

7- Arrange the pixels of the image I_1 according to the random numbers P_1 in order to obtain the permutated image I_2. The second permutation process involves arranging the pixels of the image I_2 according to the random numbers' array P_2 resulting in the image I_3.

8- For the diffusion process, generate eight chaotic sequences using multiple chaotic maps: the logistic map, the tent map and the piecewise map with eight initial conditions derived from Eq. (6). For example, three sequences using the logistic map $[X_2, X_3, X_4]$, two sequences using the tent map $[T_2, T_3]$ and three sequences using the piecewise map $[Y_1, Y_2, Y_3]$. All generated sequences are of size L.

9- Consider z as in Equation (7) as a representation for the generated maps

$$z = \{X_{2(i-L)}, T_{2(i-L)}, Y_{1(i-L)}, X_{3(i-L)}, T_{3(i-L)}, Y_{2(i-L)}, X_{4(i-L)}, Y_{3(i-L)}\} \tag{7}$$

10- Convert each element i in each sequence of z into binary representation z_{ib} using Equations (8, 9).

$$z_i = \lfloor (z_i * 10)/10 \rfloor \tag{8}$$

$$z_{ib} = \begin{cases} 0 & 0 \le z_i \le 0.5 \\ 1 & 0.5 < z_i < 1 \end{cases} \tag{9}$$

11- Concatenate z_{ib} from each corresponding sequence in z to form the binary sequence w as in Equation (10).

$$w = \{w_1, \dots, w_L\} = \{[X_{21}T_{21}Y_{11}X_{31}T_{31}Y_{21}X_{41}Y_{31}], \dots, [X_{2L}T_{2L}Y_{1L}X_{3L}T_{3L}Y_{2L}X_{4L}Y_{3L}]\} \tag{10}$$

12- For each pixel in I_3, two random numbers are generated, R_1 being in the range from one to

eight, indicates the DNA rule number, while R_2 in the range from 1 to three indicates the DNA operation. The mathematical expression used for generating pseudo random numbers is described by Equation (11).

$$R_{n+1} = (aR_n + c)\ mod\ m \tag{11}$$

The integers $a, c, and\ m$ indicate the characteristics of the random-number generator.

13- Convert each pixel in I_3 to the corresponding binary form. To diffuse the image I_3, DNA encoding process is employed based on a random rule number R_1 and one randomly selected operation from three DNA operations: XOR, XNOR and ADD using R_2.
14- Shift the value of $w(i)$ left or right according to the DNA rules. For the first four DNA rules, $w(i)$ is shifted left by a value from one to four bits, while for the second four DNA rules, $w(i)$ is shifted right.
15- Encode $I_3(i)$ and $w(i)$ using DNA rule numbers indicated by R_1. Perform one DNA operation between $I_3(i)$ and $w(i)$ according to the random value of R_2 . This value should be interpreted as follows: one indicates XOR operation, two indicates XNOR operation and three indicates Addition operation.
16- Decode the result from DNA representation to binary representation using the same rule numbers R_1 to obtain the diffused image I_4.
17- Repeat steps (8-10) to generate $\grave{w}$ another combination of a binary sequence of different eight maps using the initial conditions produced by Eq. (6).
18- Shift $\grave{w}(i)$ and $I_4(i)$ left or right according to two random numbers R_3 and R_4 generated from one to four.
19- Apply XOR function between the shifted $\grave{w}(i)$ and $I_4(i)$ to obtain the diffused image I_5. Reshape I_5 to be the cipher image.

The proposed technique is illustrated in Figure 2. By repeating the same steps in reverse order, the cipher image is decrypted to its original state.

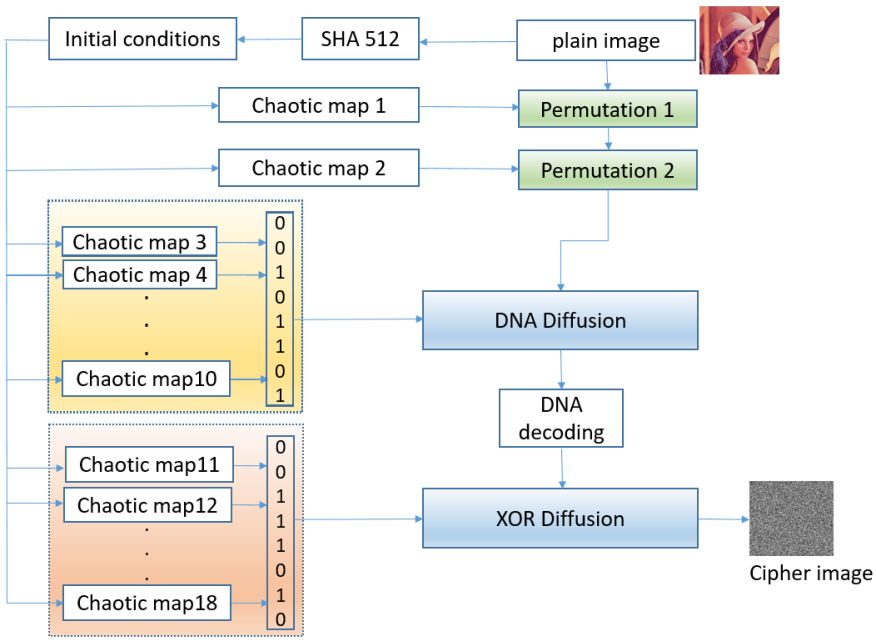


Figure 2. Summary of the proposed cryptosystem.

The proposed system is described by an example as follows.

1- Initialization:
   - Suppose an original image I_1 of size 4x4. Convert the image into 1D vector of size 1x16.

130	118	110	100
198	193	187	179
99	119	138	153
49	47	47	49

I_1	130	198	99	49	118	193	119	47	110	187	138	47	100	179	153	49

- A key is generated using the hash 256 function of the image. Let the generated key as K. All generated initial conditions are in the range [0 1], K such as

K	204	124	……	201	154

- Generate 6 chaotic sequences from each map illustrated by Eqs. (1-3) using K. The control parameter of the logistic map X is 3.8956 for the first sequence, then an increment of 0.001 is used to indicate the remaining control parameters. The same criterion is used for the tent and piecewise linear maps. For the tent map T, the control parameter of the first sequence is 1.5 with an increment of 0.05 for the following tent sequences. The control parameter of the first sequence of piecewise linear map Y is 0.25678900 with an increment of 0.0000111.
- Generate two integer random numbers, the first for DNA rule numbers in the range of [0 7] and the second for DNA operations in the range of [1 3], where one indicates XOR, two indicates XNOR and three indicates ADD.

DNA rule no. (R_1)	1	7	……	3	6
DNA operations (R_2)	2	2	……	1	3

2- Permutation:
- First permutation: sort X(1), then get the index P_1. Rearrange I_1 according to P_1 to obtain the first permutated image vector I_2.

P_1	3	10	13	16	6	4	1	8	11	14	7	5	15	12	9	2

- Second permutation: sort T(1) to obtain P_2 and I_3.

P_2	4	10	16	14	8	2	12	6	5	11	1	7	13	15	9	3

- The permutation result

Original	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
image I_1	130	198	99	49	118	193	119	47	110	187	138	47	100	179	153	49
1st perm	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
I_2	119	49	130	193	47	118	138	47	153	198	110	179	99	187	100	49
2nd perm	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
I_3	110	118	49	119	153	47	179	47	100	49	198	138	99	193	187	130

3- DNA diffusion:
- Select eight sequences from X, T and Y. Transform them into the binary form, then generate the combination W illustrated by Eq. (10). Transform pixels of I_3 into binary representation.

$I_3(1) = 110$	01101110
W = {X(2),T(2),Y(1),X(3),T(3),Y(2),Y(3),X(4)}(1)	10000100

- Perform circular shift for W, then encode the binary values into DNA representation using DNA rule number R_1.

$I_3(1)$	01101110	GCTC
Circular shift	00100001	ACAG

- Perform DNA operation R_2. Transform the obtained DNA code into binary form and obtain the diffused pixels.

XNOR	GCTC	
ACAG	CTAA	10110000
		$I_4(1) = 176$

I_3	110	118	49	119	153	47	179	47	100	49	198	138	99	193	187	130
I_4	176	181	217	43	221	91	73	69	207	206	27	97	252	112	248	135

4- XOR diffusion
   - Select eight sequences from X, T and Y. Transform them into the binary form, then generate the combination W' illustrated by Eq. (10). Transform pixels of I_4 into binary representation.

$I_4(1) = 176$	10110000
W' = {X(5),T(4),Y(4),X(6),T(5),Y(5),Y(6),T(6)}(1)	01001001

   - Apply circular shift, then XOR function to obtain the cipher image I_5.

XOR	10110000	
10100100	00010100	$I_5(1) = 20$

I_4	176	181	217	43	221	91	73	69	207	206	27	97	252	112	248	135
I_5	20	213	159	255	222	82	195	116	91	221	127	107	14	124	176	112

130	118	110	100
198	193	187	179
99	119	138	153
49	47	47	49

Original image

20	222	91	14
213	82	221	124
159	195	127	176
255	116	107	112

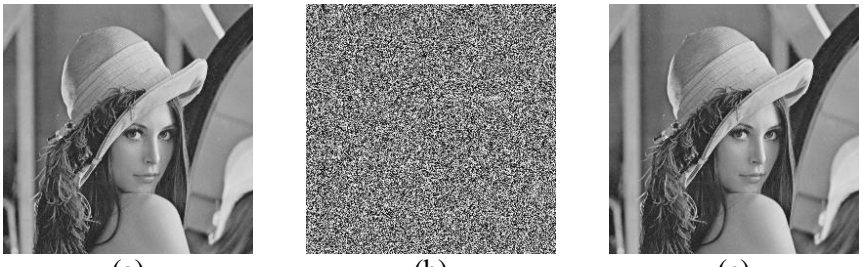
Encrypted image

5. EXPERIMENTAL RESULTS

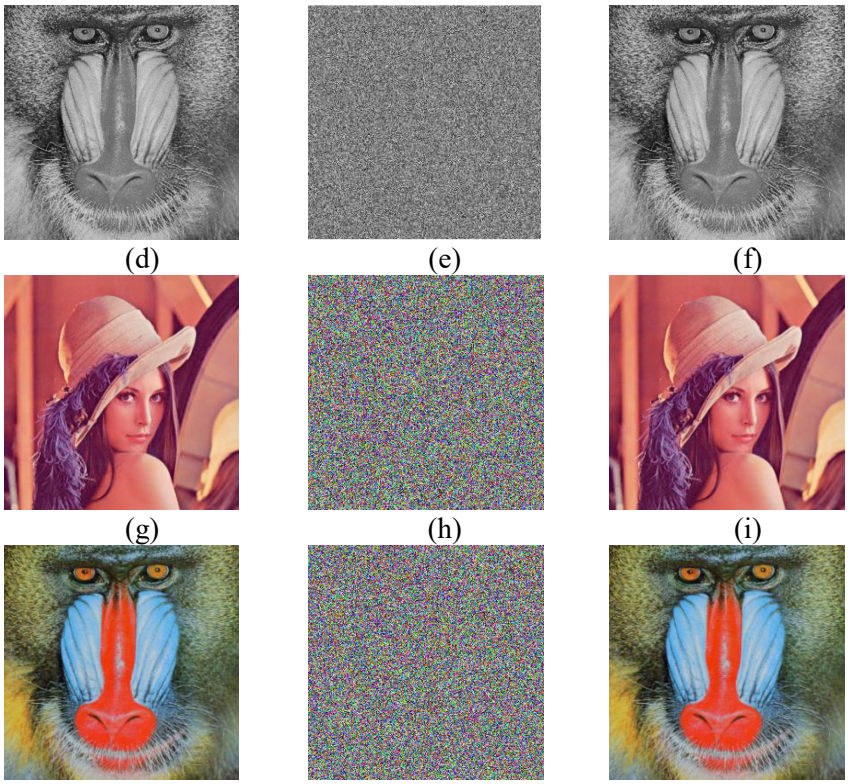
The performance of the proposed cryptosystem is evaluated using the images "Lena ", "Baboon ", "Peppers ", "Boat", "Barbara " and "Airplane ", of size 256x256 and 512x512 from the USC-SIPI image database [38]. RGB images of size 256x256x3 were also used to test the system. The simulation is performed using MATLAB R2017b on a computer with Windows 10 operating system, 4GB RAM and 3.40 GHz processor. Key generation is an important step in the encryption process. The initial values of the used chaotic maps are identified based on the generated key. There is a unique SHA 512 hash function for each original image; therefore, it is excellent to be used to generate a unique key for each image. In the proposed technique, the hash value is used to generate 18 initial values which are used to generate 18 chaotic sequences with multiple use of control parameters r=3.8956, u=1.5 and v=0.25678900. Two permutation steps are performed on the image using two chaotic maps. The keys for the diffusion process are generated by mixing the binary digits of eight chaotic maps. The first diffusion uses DNA coding along with three generated random numbers to indicate the DNA rule number, the DNA operation and the amount of circular left or right shift. The second diffusion employs an XOR operation between randomly shifted processed pixels and randomly shifted keys. Figure 3 shows original, encrypted and decrypted versions of Lena and Baboon images using the proposed technique.

5.1 Histogram Analysis

The histogram declares the distribution of pixels of an image. Figure 4 shows the histogram of gray images of Lena and Baboon before and after the encryption process. Similar results are obtained for RGB images. The difference between the histogram of the plain images and the histogram of the encrypted images is apparent. The histogram of encrypted images is flat and equally distributed, which is the opposite of a typical histogram. Therefore, the images encrypted by the proposed technique are robust against statistical attack.



(a) (b) (c)



(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

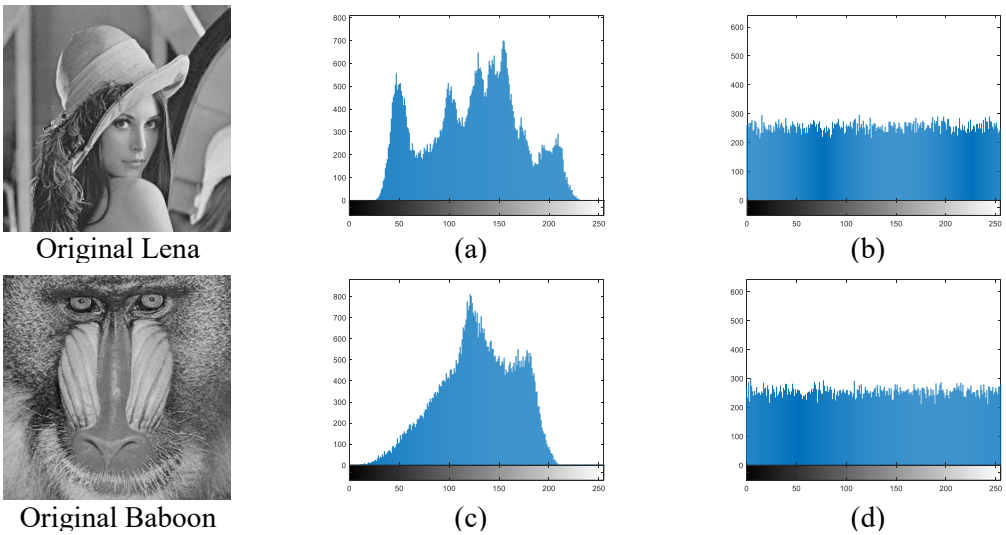
Figure 3. Plain images, cipher images and decrypted images: (a-c) gray Lena 256x256 (d-f) gray Baboon 512x512 (g-i) RGB Lena (j-l) RGB Baboon.

5.2 Chi-square Test (X^2)

X^2 test is a test that measures how well a model fits the observed data. In the case of encryption, the distribution of pixels in an image can be measured using X^2. The theoretical value of X^2 for the encrypted image is 293.24783 [39]. Any encrypted image should have an X^2 value lower than the theoretical value. The value of X^2 can be calculated using Equation (12).

$$X^2 = \sum_i^{256} (p_i - 256)^2/256 \quad (12)$$

where i is the gray scale levels and p_i is the corresponding frequency occurrences. Tables 3 and 4 display the X^2 values for the plain images Lena, Baboon, Boat, Peppers, Barbara and Airplane, as well as their respective X^2 values after encryption. It is evident that all X^2 values of encrypted images are lower than the theoretical values.



Original Lena (a) (b)

Original Baboon (c) (d)

Figure 4. Histogram analysis: (a, b) histogram of original and encrypted Lena 256x256, (c-d) histogram of original and encrypted Baboon 512x512.

Table 3. Chi-square values of plain images (256x256) and the corresponding cipher images.

Images	Lena	Baboon	Boat	Peppers	Airplane	Barbara
Plain	39868.726	58107.609	100674.875	33094.195	174458.335	27793.562
Encrypted	243.3984	280.1857	246.8359	216.6406	239.1250	290.2109
[29]	252.7891	261.3301	282.5039	-	-	-
[3]	286.4766	219.5625	-	214.7813	-	-
[21]	222.0156	247.9766	-	237.375	-	-

Table 4. Chi-square values of RGB plain images (256x256x3) and the corresponding cipher images.

Images	Lena	Baboon	Boat	Peppers	Airplane	Barbara
Plain	65305.64	29129.86	54151.70	57105.97	167182.92	28923.28
Encrypted	312.0391	268.7500	238.2422	199.8281	237.7109	236.9219

5.3 Entropy Analysis

Entropy measures the randomness of the data. According to images, it shows the distribution of pixels in an image. The maximum value of entropy is eight, where all the pixels are uniformly distributed with 1⁄256 probability of occurrence for each. Therefore, an encrypted image achieve randomness if its entropy is close to eight. The entropy is computed using Equation (13).

$$E(m) = \sum_{i=0}^{2^n-1} p(m_i) \log_2 \frac{1}{p(m_i)} \tag{13}$$

where n is the number of bits used to represent a pixel and $p(m_i)$ is the probability of occurrence of a pixel (m_i). Tables 5 and 6 show entropy values of the proposed technique using standard images. It is clear that all measured entropy values are close to eight approximately above 7.997 for 256x256 images and 7.9992 for 512x512 images, which ensures the randomness of the proposed technique.

Table 5. Entropy of the proposed technique using standard images of size 256x256.

Images	Plain Image	Cipher Image	[3]	[23]	[20]
Lena	7.4429	7.9979	7.9969	7.9972	7.9894
Peppers	7.5620	7.9976	7.9976	-	-
Airplane	6.7261	7.9974	-	-	-
Baboon	7.2374	7.9972	7.9976	-	-
Boat	7.1587	7.9973	-	-	-

Table 6. Entropy of the proposed technique using standard images of size 512x512.

Images	Proposed technique	[29]	[31]	[20]
Lena	7.9994	7.99935	-	7.9916
Peppers	7.9992	7.99921	7.9993	-
Airplane	7.9993	-	-	7.9916
Baboon	7.9992	7.99928	7.9993	7.9914
Boat	7.9993	7.99921	-	7.9916

5.4 Correlation Analysis

Correlation analysis is used to measure the similarity between the original image and the encrypted image. In the original image, the correlation between adjacent pixels is approximately one because of high similarity between adjacent pixels. In the encrypted image, the similarity between adjacent pixels should be none; therefore, the correlation between them should be zero. The correlation between adjacent pixels can be calculated using Equations (14-17) [38].

$$\bar{x} = 1/N \sum_{i=1}^{N} x_i \tag{14}$$

$$D(x) = 1/N \sum_{i=1}^{N} (x_i - \bar{x})^2 \tag{15}$$

$$cov(x, y) = 1/N \sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y}) \tag{16}$$

$$C_r = cov(x, y)/\sqrt{D(x)D(y)} \quad (17)$$

where N is the total number of pixels $m * n$. To calculate the correlation between adjacent pixels for both original image and encrypted image, two arrays x, y are created of 3000 randomly selected pixels and their corresponding adjacent pixels for the three directions horizontal, vertical and diagonal. Table 7 shows the correlation values between adjacent pixels for both original and encrypted standard gray images Lena, Peppers, Airplane, Baboon, Boat and Barbara. It is clear that the correlation value of adjacent pixels of encrypted images obtained using the proposed technique is significantly reduced, where the maximum value is 0.007. Table 8 shows the correlation of adjacent pixels for both original and encrypted standard RGB images Lena, Peppers, Airplane, Baboon, Boat and Barbara. As we can see, the correlation between adjacent pixels for the three colour channels of the encrypted images is very low. The introduced results are emphasized by Figure 5, which depicts the correlation of plain and encrypted Lena images in the three directions horizontal, vertical and diagonal. These results indicate the security and resistance of the proposed algorithm.

Table 7. Correlation analysis of the gray images using the proposed technique.

Gray Image	***Plain Image***			***Cipher Image***		
	Horizontal	**Vertical**	**Diagonal**	**Horizontal**	**Vertical**	**Diagonal**
Lena	0.9258	0.9593	0.9037	-0.0029	0.0019	0.0076
Baboon	0.8700	0.8410	0.7888	0.00003	-0.00003	-0.0082
Airplane	0.9396	0.9331	0.8883	0.0010	0.0012	-0.0039
Peppers	0.9675	0.9729	0.9432	0.0040	0.0034	0.0085
Boat	0.9268	0.9452	0.8833	0.0081	-0.0024	0.0039
Barbara	0.9342	0.9567	0.9056	-0.0072	-0.0016	0.0031

Table 8. Correlation analysis of the RGB Images using the proposed technique.

Image	**Color channels**	**Original Image**			**Encrypted Image**		
		H	**V**	**D**	**H**	**V**	**D**
Lena	R	0.9558	0.9781	0.9336	-0.0034	0.0053	0.0030
	G	0.9401	0.9695	0.9180	-0.00009	-0.0010	0.00007
	B	0.9189	0.9495	0.8948	0.0037	0.00005	0.0029
Peppers	R	0.9646	0.9680	0.9369	0.0064	0.0012	0.0027
	G	0.9698	0.9750	0.9466	0.00002	0.0043	0.0051
	B	0.9570	0.9636	0.9263	-0.0004	-0.0038	0.0044
Airplane	R	0.9389	0.9239	0.8738	-0.0040	-0.0019	0.0053
	G	0.9309	0.9343	0.8814	-0.0007	-0.0054	-0.0036
	B	0.9503	0.9089	0.8800	0.0076	-0.0030	-0.0013
Baboon	R	0.9105	0.8595	0.8474	0.0035	-0.0017	0.0028
	G	0.8594	0.7755	0.7434	0.0011	-0.0012	-0.0004
	B	0.8953	0.8697	0.8296	-0.0003	0.0050	-0.0063
Boat	R	0.9563	0.9539	0.9274	0.0038	-0.0054	-0.0060
	G	0.9558	0.9527	0.9225	0.0019	0.0032	-0.0034
	B	0.9603	0.9645	0.9369	-0.0003	0.0033	-0.0001
Barbara	R	0.9526	0.9611	0.9182	0.0038	-0.00003	0.0005
	G	0.9445	0.9543	0.9029	0.0007	-0.0046	-0.0027
	B	0.9526	0.9624	0.9182	0.0014	0.0015	-0.0004

5.5 Plaintext Sensitivity Analysis

A cryptosystem must be robust against any slight change in the plaintext image. Therefore, if a randomly selected pixel of the plaintext image is slightly changed, the resulting encrypted image will be different. Two metrics are used to evaluate the plaintext sensitivity of a cryptosystem: NPCR (number of pixel changing rate) and UACI (unified averaged changed intensity) [40]. The mathematical representations of NPCR and UACI are given by Equations (18) and (19).

$$NPCR(E_1,E_2) = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N} D(i,j)=\begin{cases}1, & if\ E_1(i,j)\neq E_2(i,j)\\ 0, & if\ E_1(i,j)=E_2(i,j)\end{cases}}{M\times N} \times 100\% \quad (18)$$

$$UACI(E_1,E_2) = \frac{1}{M\times N}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{|E_1(i,j)-E_2(i,j)|}{255} \times 100\% \quad (19)$$

where E_1 and E_2 are two encrypted images corresponding to two plaintext images differing by a slight difference in only one randomly selected pixel. M and N are the dimensions of the image. The cryptosystem is robust and sensitive to plaintext image if the value of NPCR is greater than 99% and the value of UACI is around 33% [3]. Table 9 shows the achieved values of NPCR and UACI for the images Lena, Pepper, Airplane, Baboon, Boat and Barbara. The results demonstrate the robustness of the proposed cryptosystem against differential attacks.

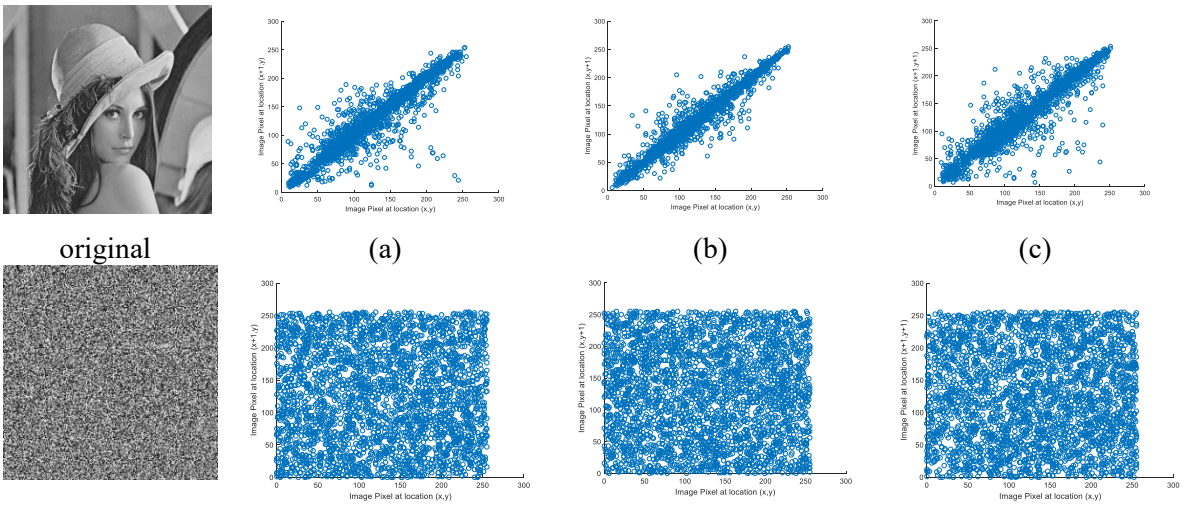


Figure 5. Correlation of adjacent pixels of plain and encrypted Lena images: (a, d) horizontal direction, (b, e) vertical direction, (c, f) diagonal direction.

Table 9. NPCR and UACI values for grayscale images of size 256x256.

Image	Selected pixel	NPCR(%)	UACI(%)	[3]
Lena	P(1,1)	99.63	33.46	99.64/33.42
Peppers	P(167,233)	99.61	33.50	99.64/33.40
Airplane	P(217,17)	99.64	33.57	-
Baboon	P(34,50)	99.68	33.61	99.64/33.43
Boat	P(128,128)	99.63	33.45	-
Barbara	P(256,256)	99.68	33.47	-

5.6 Speed Analysis

An important issue of the cryptosystem is the time efficiency. The speed of the proposed technique is evaluated based on Intel Core™i7 CPU at 2.60 GHz computer using Matlab 2017b. The proposed technique consumes about three seconds on average to encrypt an image of size 256x256.

5.7 Key Space Analysis

The key space is the total number of possible keys used in a cryptosystem. To be secure against brute-force attacks, the key space must be so large that it is infeasible for an intruder to compute the key and decrypt the image. The encryption algorithm is robust if the key space is larger than 2^{100} [3]. In the proposed technique, the key is formed from 18 chaotic maps with 18 initial conditions and 18 control parameters. If each variable has a computational accuracy of 10^{-15}, then the provided key space is 10^{540}. This large key space indicates that the proposed technique is highly resistant to brute-force attacks.

5.8 Key Sensitivity Test

The encryption system is key sensitive if the encrypted image is completely changed according to any slight change in the key. Therefore, the decryption process should fail if the original key is altered by small margin. To test the sensitivity of the proposed system, one initial condition is changed by adding 10^{-15} to its original value. The sensitivity-test results are shown in Figure 6. It is shown that the decryption of the cipher image failed due to the minor alteration that was introduced.

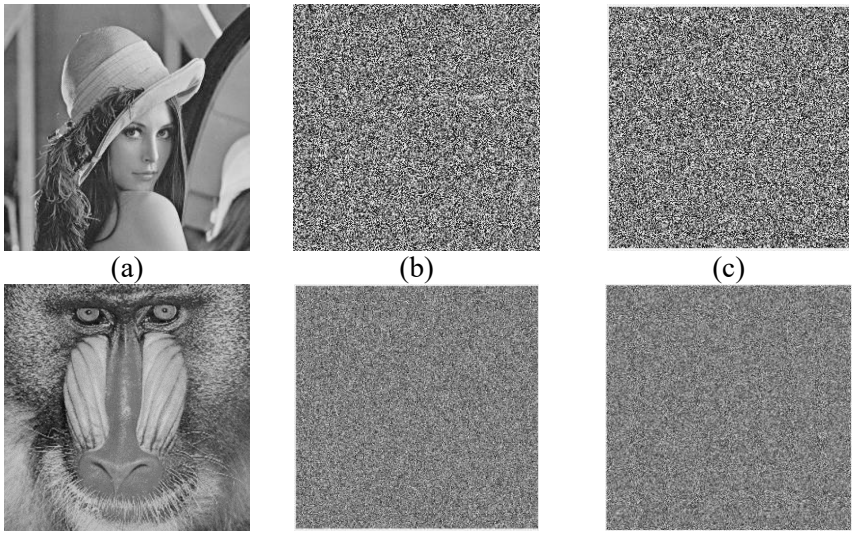


Figure 6. Key sensitivity test: (a, d) original images of Lena and Baboon, (b, e) encrypted images, (c, f) decrypted images.

5.9 Robustness Test

Images that are conveyed *via* communication channels may experience degradation due to the presence of noise introduced by either a malicious attacker or inherent characteristics of the channel, resulting in a decline in the overall quality of the encrypted images. The decrypted images' quality is influenced to some extent by noisy pixels; hence, the encryption method should possess a degree of resilience against noise interference. The robustness of the proposed technique is tested by adding 10% and 30% of salt and pepper noise to the original images, then encrypting the noisy images. Figures 7 and 8 show the encrypted and decrypted images subjected to salt and pepper noise. The figures illustrate the efficacy of the proposed system in effectively handling noisy images and extracting crucial data to the fullest extent.

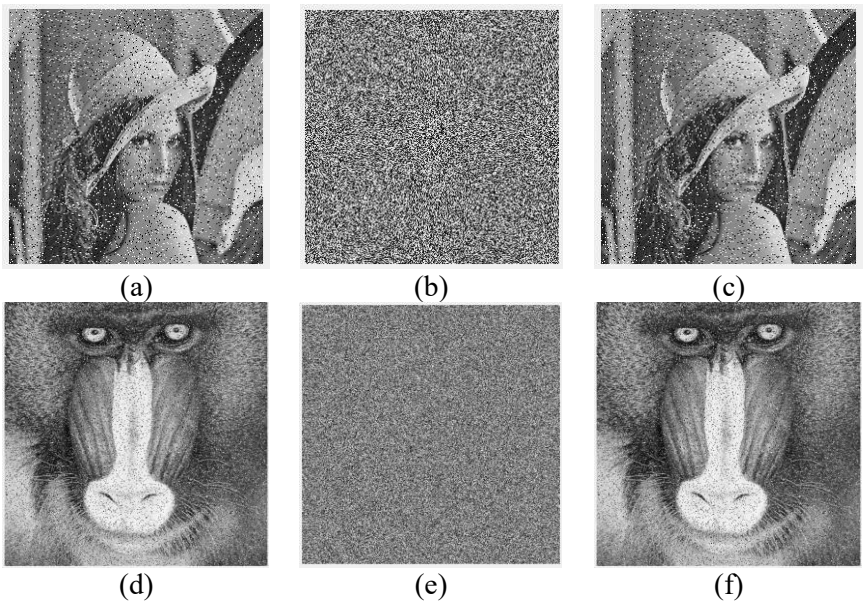


Figure 7. Robustness test (10% salt and pepper noise): (a, d) noisy original images of Lena and Baboon, (b, e) encrypted images, (c, f) decrypted images.

Table 10. Statistical test comparison for Lena images of size 256x256.

Metric	*Entropy*	*Chi-square*	*NPCR (%)*	*UACI (%)*	*Correlation Analysis*		
					Horizontal	**Vertical**	**Diagonal**
Proposed	7.9979	243.3984	99.63	33.46	-0.0029	0.0019	0.0076
[3]	7.9969	286.4766	99.64	33.42	0.0222	0.0354	0.0006
[20]	7.9894	-	99.66	33.42	0.0044	0.0015	0.0019
[41]	7.9992	-	99.614	33.364	0.0019	0.0014	0.0052
[21]	7.9976	222.0156	99.61	33.51	0.0305	−0.0043	0.0042
[23]	7.9972	-	99.63	-	-0.0028	-0.00006	-0.0011

Table 11. Statistical test comparison for Baboon images of size 512x512.

Metric	*Entropy*	*Chi-square*	*NPCR (%)*	*UACI (%)*	*Correlation Analysis*		
					Horizontal	**Vertical**	**Diagonal**
Proposed	7.9992	302.7969	99.61	33.45	4.644e-04	2.42e-04	8.622e-04
[41]	7.9998	-	99.645	33. 426	0.0024	0.0054	0.0041
[22]	7.9992	-	0.9959	0.3352	0.9671	0.9744	0.9381
[42]	7.9992	-	0.9960	0.3349	0.7508	0.8562	0.7153

The outcomes of the statistical and differential analyses of the proposed technique are compared with those of a few schemes in the literature for the images of Lena and Baboon as illustrated in Tables 10 and 11, respectively. The comparison evinces the efficacy and resilience of the proposed technique.

6. CONCLUSIONS

This paper proposes a new encryption technique based on the multiple use of different one-dimensional chaotic maps. Two chaotic sequences are used to scramble the pixel positions. The key stream for the diffusion process is composed of the concatenation of the corresponding binary digits of the generated random sequences. Each generated key is circularly rotated by a random value. DNA coding is applied based on three different DNA functions (XOR, XNOR and Addition). Random-number generators are used to select the DNA function, the DNA rule number and the value of the circular rotation. The proposed technique is applied to standard grayscale images and RGB images. The simulation results demonstrate the effectiveness and high security of the proposed technique.

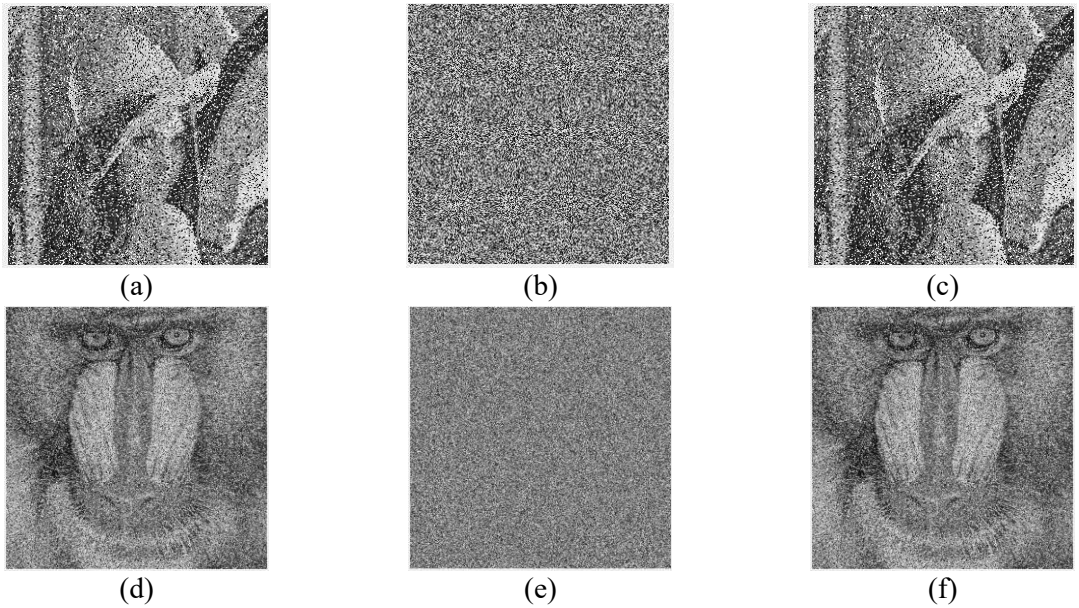


Figure 8. Robustness test (30% salt and pepper noise): (a, d) noisy original images of Lena and Baboon, (b, e) encrypted images, (c, f) decrypted images.

CONFLICT OF INTERESTS

The author declares no conflict of interests.

REFERENCES

[1] N.-R. Zhou, L.-J. Tong and W.-P. Zou, "Multi-image Encryption Scheme with Quaternion Discrete Fractional Tchebyshev Moment Transform and Cross-coupling Operation," Signal Processing, vol. 211, p. 109107, 2023.

[2] E. Yavuz, "A Novel Chaotic Image Encryption Algorithm Based on Content-sensitive Dynamic Function Switching Scheme," Optics & Laser Technology, vol. 114, pp. 224-239, 2019.

[3] M. Akraam, T. Rashid and S. Zafar, "A Chaos-based Image Encryption Scheme is Proposed Using Multiple Chaotic Maps," Mathematical Problems in Engineering, vol. 2023, DOI: 10.1155/2023/2003724, 2023.

[4] J. Fridrich, Steganography in Digital Media: Principles, Algorithms and Applications, DOI: 10.1017/CBO9781139192903, Cambridge University Press, 2009.

[5] J. Buchmann, Introduction to Cryptography, 2nd Edn, ISBN: 0387207562, Springer, 2004.

[6] S. Toughi, M. H. Fathi and Y. A. Sekhavat, "An Image Encryption Scheme Based on Elliptic Curve Pseudo Random and Advanced Encryption System," Signal Processing, vol. 141, pp. 217-227, 2017.

[7] E. Tromer, D. A. Osvik and A. Shamir, "Efficient Cache Attacks on AES and Countermeasures," Journal of Cryptology, vol. 23, pp. 37-71, 2010.

[8] R. Shivhare, R. Shrivastava and C. Gupta, "An Enhanced Image Encryption Technique Using DES Algorithm with Random Image Overlapping and Random Key Generation," Proc. of the 2018 IEEE Int. Conf. on Advanced Computation and Telecommunication (ICACAT), pp. 1-9, Bhopal, India, 2018.

[9] Q. Zhang and Q. Ding, "Digital Image Encryption Based on Advanced Encryption Standard (AES)," Proc. of the 2015 5th IEEE Int. Conf. on Instrumentation and Measurement, Computer, Communication and Control (IMCCC), pp. 1218-1221, Qinhuangdao, China, 2015.

[10] H. Kolivand, S. F. Hamood, S. Asadianfam and M. S. Rahim, "Image Encryption Techniques: A Comprehensive Review," Multimedia Tools and Applications, pp. 1-36, DOI: 10.1007/s11042-023-17896-0, 2024.

[11] U. Zia et al., "Survey on Image Encryption Techniques Using Chaotic Maps in Spatial, Transform and Spatiotemporal Domains," Int. Journal of Information Security, vol. 21, pp. 917-935, 2022.

[12] R. Matthews, "On the Derivation of a "Chaotic" Encryption Algorithm," Cryptologia, vol. 13, pp. 29-42, 1989.

[13] B. Yang and X. Liao, "A New Color Image Encryption Scheme Based on Logistic Map over the Finite Field ZN," Multimedia Tools and Applications, vol. 77, pp. 21803-21821, 2018.

[14] L. Meng, S. Yin, C. Zhao, H. Li and Y. Sun, "An Improved Image Encryption Algorithm Based on Chaotic Mapping and Discrete Wavelet Transform Domain," Int. Journal of Network Security, vol. 22, pp. 155-160, 2020.

[15] B. Mondal, S. Singh and P. Kumar, "A Secure Image Encryption Scheme Based on Cellular Automata and Chaotic Skew Tent Map," Journal of Information Security and Applications, vol. 45, pp. 117-130, 2019.

[16] D. S. Malik and T. Shah, "Color Multiple Image Encryption Scheme Based on 3D-Chaotic Maps," Mathematics and Computers in Simulation, vol. 178, pp. 646-666, 2020.

[17] Y. Qobbi, A. Jarjar, M. Essaid and A. Benazzi, "Image Encryption Algorithm Using Dynamic Permutation and Large Chaotic S-Box," Multimedia Tools and Appli., vol. 82, pp. 18545-18564, 2023.

[18] V. Kumar and A. Girdhar, "A 2D Logistic Map and Lorenz-Rossler Chaotic System Based RGB Image Encryption Approach," Multimedia Tools and Applications, vol. 80, pp. 3749-3773, 2021.

[19] B. Rahul, K. Kuppusamy and A. Senthilrajan, "Dynamic DNA Cryptography-based Image Encryption Scheme Using Multiple Chaotic Maps and SHA-256 Hash Function," Optik, vol. 289, p. 171253, 2023.

[20] T. Li, B. Du and X. Liang, "Image Encryption Algorithm Based on Logistic and Two-dimensional Lorenz," IEEE Access, vol. 8, pp. 13792-13805, 2020.

[21] Y. Niu, Z. Zhou and X. Zhang, "An Image Encryption Approach Based on Chaotic Maps and Genetic Operations," Multimedia Tools and Applications, vol. 79, pp. 25613-25633, 2020.

[22] J. Wu, X. Liao and B. Yang, "Image Encryption Using 2D Hénon-Sine Map and DNA Approach," Signal Processing, vol. 153, pp. 11-23, 2018.

[23] U. K. Gera and S. Agrawal, "Image Encryption Using a Combination of 4D Discrete Hyperchaotic Map and DNA Encoding," Multimedia Tools and Applications, vol. 83, pp. 38037-38054, 2023.

[24] Y. Wan, S. Gu and B. Du, "A New Image Encryption Algorithm Based on Composite Chaos and Hyperchaos Combined with DNA Coding," Entropy, vol. 22, p. 171, 2020.

[25] R. Enayatifar, A. H. Abdullah and I. F. Isnin, "Chaos-based Image Encryption Using a Hybrid Genetic Algorithm and a DNA Sequence," Optics and Lasers in Engineering, vol. 56, pp. 83-93, 2014.

[26] H. Liu, L. Teng, Y. Zhang, R. Si and P. Liu, "Mutil-medical Image Encryption by a New Spatiotemporal Chaos Model and DNA New Computing for Information Security," Expert Systems with Applications, vol. 235, p. 121090, 2024.
[27] Q. Lai, L. Yang, G. Hu, Z.-H. Guan and H. H.-C. Iu, "Constructing Multiscroll Memristive Neural Network with Local Activity Memristor and Application in Image Encryption," IEEE Transactions on Cybernetics, vol. 54, no. 7, pp. 4039-4048, 2024.
[28] Q. Lai, L. Yang and G. Chen, "Design and Performance Analysis of Discrete Memristive Hyperchaotic Systems with Stuffed Cube Attractors and Ultraboosting Behaviors," IEEE Transactions on Industrial Electronics, vol. 71, no. 7, pp. 7819-7828, 2023.
[29] P. Liu, X. Wang, X. Zhao and S. Unar, "Target-based Image Encryption *via* Infinite Interval Chaotic System with Ill-conditioned Parameter and 3DBDM," Expert Systems with Applications, vol. 232, p. 120811, 2023.
[30] P. Liu, X. Wang and Y. Su, "Image Encryption *via* Complementary Embedding Algorithm and New Spatiotemporal Chaotic System," IEEE Transactions on Circuits and Systems for Video Technology, vol. 33, pp. 2506-2519, 2022.
[31] X. Wang and P. Liu, "A New Full Chaos Coupled Mapping Lattice and Its Application in Privacy Image Encryption," IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 69, pp. 1291-1301, 2021.
[32] P. Liu, X. Wang, Y. Su, H. Liu and S. Unar, "Globally Coupled Private Image Encryption Algorithm Based on Infinite Interval Spatiotemporal Chaotic System," IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 70, pp. 2511-2522, 2023.
[33] Q. Lai and Y. Liu, "A Cross-channel Color Image Encryption Algorithm Using Two-dimensional Hyperchaotic Map," Expert Systems with Applications, vol. 223, p. 119923, 2023.
[34] R. B. Naik and U. Singh, "A Review on Applications of Chaotic Maps in Pseudo-random Number Generators and Encryption," Annals of Data Science, vol. 11, pp. 25-50, 2024.
[35] C. Li, G. Luo, K. Qin and C. Li, "An Image Encryption Scheme Based on Chaotic Tent Map," Nonlinear Dynamics, vol. 87, pp. 127-133, 2017.
[36] X. Wang and C. Jin, "Image Encryption Using Game of Life Permutation and PWLCM Chaotic System," Optics Communications, vol. 285, pp. 412-417, 2012.
[37] I. I. Cisse, H. Kim and T. Ha, "A Rule of Seven in Watson-Crick Base-pairing of Mismatched Sequences," Nature Structural & Molecular Biology, vol. 19, pp. 623-627, 2012.
[38] R. Enayatifar, F. G. Guimarães and P. Siarry, "Index-based Permutation-diffusion in Multiple-image Encryption Using DNA Sequence," Optics and Lasers in Engineering, vol. 115, pp. 131-140, 2019.
[39] L. Liu, Y. Lei and D. Wang, "A Fast Chaotic Image Encryption Scheme with Simultaneous Permutation-Diffusion Operation," IEEE Access, vol. 8, pp. 27361-27374, 2020.
[40] Y. Wu, J. P. Noonan and S. Agaian, "NPCR and UACI Randomness Tests for Image Encryption," Cyber Journals: Multidisciplinary Journals in Science and Technology, Journal of Selected Areas in Telecommunications (JSAT), vol. 1, pp. 31-38, 2011.
[41] Y. Qobbi, A. Jarjar, M. Essaid and A. Benazzi, "Image Encryption Algorithm Based on Genetic Operations and Chaotic DNA Encoding," Soft Computing, vol. 26, pp. 5823-5832, 2022.
[42] X. Chai, Z. Gan, K. Yang, Y. Chen and X. Liu, "An Image Encryption Algorithm Based on the Memristive Hyperchaotic System, Cellular Automata and DNA Sequence Operations," Signal Processing: Image Communication, vol. 52, pp. 6-19, 2017.

ملخص البحث:

تقدّم هذه الورقة تقنيةً لتشفير الصُّور بناءً على خرائط فوضى متعدّدة أحادية البُعد مع ترميز DNA. وتستخدم الطريقة المقترحة ثلاث خرائط فوضى أحادية البُعد، هي الخريطة اللوجستية، وخريطة الخيمة، وخريطة "قطعة - قطعة" عدّة مرّات لتوليد ثمانية عشر تتابعاً عشوائياً بقيم ابتدائية مُختلفة ومتغيرات مُختلفة. ولتشفير الصُّور، تُستخدم العناصر الثّنائية من التّتابُعات المتنوّعة لخرائط الفوضى لتغيير شدّة إضاءة النُّقَط في الصّورة المزمع تشْفيرها في عملية الانتشار. ويستخدم ترميز DNA من خلال الاختيار العشوائي للقواعد والعمليات الخاصّة بكلّ نقطة من نُقَط الصّورة.

وقد جرى تقييم التّقنية المقترحة باستخدام العديد من الصُّور المعيارية، وذلك بناءً على قياس عددٍ من معايير الأداء. وقد برهنت نتائج التّقييم على توفُّر أمْن المعلومات في الطّريقة المقترحة وقدرتها على مقاومة الهجمات.

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 10, No. 04, December 2024.

ENHANCING DIAGNOSTIC ACCURACY WITH ENSEMBLE TECHNIQUES: DETECTING COVID-19 AND PNEUMONIA ON CHEST X-RAY IMAGES

Fatma A. Mostafa[1], Lamiaa A. Elrefaei[1], Mostafa M. Fouda[2] and Aya Hossam[1]

(Received: 27-May-2024, Revised: 17-Jul.-2024, Accepted: 12-Aug.-2024)

ABSTRACT

Lung diseases such as COVID-19 and pneumonia can lead to severe complications, including breathing difficulties, decreased lung function and respiratory failure, which can be life-threatening if not promptly treated. Chest X-ray imaging techniques have proven to be quick, effective and cost-efficient in diagnosing and monitoring these diseases. Additionally, artificial intelligence, particularly through deep learning and machine learning, has shown promising results in detecting various lung diseases, including COVID-19 and pneumonia. This technology's ability to analyze large datasets rapidly has contributed to reducing the spread of these diseases and has significantly advanced biomedical research in various medical disciplines. In this research paper, we introduced various advanced ensemble techniques as bagging, boosting, stacking and blending with different algorithms, to enhance the performance of our classification models in detecting coronavirus and pneumonia. We specifically focused on combining convolutional neural network (CNN) and vision transformer (ViT) models to create powerful ensemble models. Our objective was to determine the most accurate ensemble technique for diagnosing lung diseases. We assessed their ability to correctly classify chest X-ray images as either COVID-19, pneumonia or normal. The CatBoost model achieved the highest accuracy, F1-score and ROC-AUC score of 99.753%, 99.51% and 99.99%, respectively using the COVID-19 Radiography dataset. The bagging ensemble model achieved the highest accuracy, F1-score and ROC-AUC score of 95.08%, 95.2% and 99.69%, respectively using COVIDx CXR-4. The results indicate that the advanced ensemble techniques can significantly improve the performance of machine-learning models.

KEYWORDS

COVID-19, CXR, DenseNet-169, ViT-l32, Stacking, Boosting, Bagging, Blending.

1. INTRODUCTION

Lung diseases are a group of conditions that affect the health of the respiratory system and include many different diseases, such as pneumonia, COVID-19, pulmonary fibrosis and asthma [1]. One of the diseases that currently poses a major challenge is the new Coronavirus (COVID-19), where the number of deaths exceeded 7 million cases until April 2024 [2]. Coronavirus cases are currently a common lung disease. The new coronavirus mutates rapidly, leading to an increase in infection cases, with the number of infections reaching more than 700 million cases [2]. Pneumonia is not less dangerous than COVID-19. Therefore, early detection of these diseases reduces their risk [3]. One of the methods that is widely used in the detection of lung diseases is X-rays, as it is fast and inexpensive compared to other methods [4]. While X-rays are valuable in the detection of lung diseases, they may pose potential health risks with repeated exposure to ionizing radiation. In addition, while X-rays provide valuable information about the structure of the lungs, they may not always provide detailed insights into specific lung conditions, such as distinguishing between different types of pneumonia or identifying lung cancer at an early stage. In such cases, additional imaging techniques, such as computed tomography (CT), magnetic resonance imaging (MRI) or diagnostic tests, may be necessary for a comprehensive evaluation.

Artificial intelligence (AI) has revolutionized the medical field by searching medical data and revealing insights to enhance patient experiences and health outcomes [5]. It does this by utilizing machine-learning and deep-learning models. AI is frequently used in medical-imaging settings, analyzing CT scans, X-rays, MRIs and other images to look for lesions or other findings that a human radiologist might overlook, in addition to dealing with a huge volume of data quickly. Chest X-rays

1. F. A. Mostafa, L. A. Elrefaei and A. Hossam are with Department of Electrical Engineering, Faculty of Engineering at Shoubra, Benha University, Cairo, Egypt. Emails: fatma.mostafa13@feng.bu.edu.eg, lamia.alrefaai@feng.bu.edu.eg and aya.ahmed@feng.bu.edu.eg
2. Mostafa M. Fouda is with Department of Electrical and Computer Engineering, Idaho State University, Pocatello, ID, USA. Email: mfouda@ieee.org

(CXR) help detect lung abnormalities early and are also less expensive than other tests [6].

There are many types of deep-learning algorithms, such as CNNs, Multilayer Perceptrons (MLPs), Recurrent Neural Networks (RNNs) and Autoencoders, which are mainly used for image classification and object detection [7]-[9].

Ensemble methods are techniques the goal of which is to improve the performance of machine-learning models by creating one more reliable model by combining multiple models. Through the use of these techniques, a more reliable and more accurate model is produced by merging the predictions of several separate models. The main types of ensemble techniques are: bagging, boosting, stacking and blending techniques [10]. The bagging ensemble technique is mainly applied in classification which can increase accuracy and eliminate overfitting [11]. Boosting ensemble technique combines several weak learners to form a strong one; it takes many algorithms including gradient boosting, adaptive boosting (AdaBoost), light gradient-boosting machine (LightGBM) and extreme gradient boosting (XGBoost) [12]. The stacking ensemble technique uses predictions for multiple models to build a new strong model [13]. The blending ensemble technique also combines predictions from multiple models as stacking and can improve the overall performance of the model [14]-[15].

The integration of AI and deep-learning algorithms with ensemble methods holds significant promise for advancing the capabilities of medical imaging, leading to more accurate and more efficient detection and diagnosis of lung diseases and other medical conditions. We must take several considerations associated with the integration of AI and deep-learning algorithms with ensemble methods in medical imaging as data-management considerations, including storage, retrieval and processing, as well as algorithmic methods for disease classification and segmentation to ensure the successful implementation of these technologies in clinical practice.

The key contributions of this paper are:

1. Involving a comprehensive evaluation of the optimized CNN and ViT models.
2. Optimizing CNN and ViT hyper-parameters: The paper focuses on optimizing the hyper-parameters of Convolutional Neural Network (CNN) and Vision Transformer (ViT) models to reduce model losses and achieve the best accuracy in diagnosing lung diseases, particularly pneumonia and COVID-19.
3. Application of different ensemble techniques between CNN and ViT: The study explores the application of different ensemble techniques between CNN and ViT models. This involves leveraging ensemble methods to combine the strengths of these two architectures for improved diagnostic accuracy in lung-disease classification.
4. Comparison between the types of advanced ensemble techniques and fusion: The paper makes a comparison between different types of advanced ensemble techniques and fusion methods. This comparison provides insights into the effectiveness of various ensemble approaches in enhancing the accuracy of lung-disease diagnosis.
5. Use of the upgraded ensemble model to classify and recognize lung diseases: The upgraded ensemble model is utilized to classify and recognize lung diseases, such as pneumonia and COVID-19. This involves leveraging the optimized CNN and ViT models, along with ensemble techniques, to achieve accurate classification and recognition of lung diseases based on medical-imaging data.

This paper is organized as follows: Section 2 highlights related work. Next, Section 3 discusses the proposed work, Section 4 discusses the experimental results and lastly, conclusions and future-research directions are highlighted in Section 5.

2. RELATED WORK

We divided the previous studies according to the types of ensemble techniques that can be used to diagnose lung diseases based on X-ray images, including methods, such as bagging, stacking, boosting, blending and weighted-average techniques.

Hasan et al. [16] created a model for automatically detecting pneumonia using CXR images. The weighted-average ensemble model was applied to three models; namely, VGG16, MobileNetV2 and DenseNet169. The ensemble model achieved an accuracy of 92%. Tang et al. [17] also presented a

weighted-average ensemble approach to detect COVID-19 with an accuracy of 95%. They used COVIDx dataset to evaluate their experiment. They used COVID-Net as the candidate to generate multiple model snapshots in terms of its promising performance for its CXR image-based COVID-19 case detection.

Govindarajan et al. [18] presented an Extreme Gradient Boosting (XGBoost) classifier that can detect tuberculosis (TB) disease using CXR images from NIAID TB Portals repository. All images were 401 at a resolution of 1024×1024 pixels. The accuracy resulting from this model was 93%. Also, Kalaivani et al. [19] presented an ensemble boosted model using CNN and four different classifiers (decision tree, random forest, AdaBoost and support vector machine) to detect COVID-19. The suggested model achieved an accuracy of 99.35%. The images used were 5178 abnormal CXR images and 4310 normal CXR images collected from different sources.

Soundrapandiyan et al. [20] introduced a stacked ensemble model for detecting coronavirus (COVID-19) from chest X-ray images. The stacked ensemble technique between the models ResNet50, VGG19, Xception and DarkNet19 is named WavStaCovNet-19. The model achieved an accuracy of 94.24% on 4 classes (COVID-19, viral pneumonia, bacterial pneumonia and normal). They used two datasets, the COVID-19 image data-collection repository and chest X-ray images for normal cases and pneumonia [21]. Huang et al. [22] also presented a stacking ensemble model on the classification of multiple chest diseases including COVID-19 using the COVID-19 Radiography dataset. The model achieved an accuracy of 99.21%, Precision of 99.23%, Recall of 99.25%, F1-score of 99.20% and (area under the curve) AUC of 99.51% on the chest X-ray dataset. They verified that the combined model had better performance than individual pre-trained models. Six EfficientNetV2 models including EfficientNetV2-B0, EfficientNetV2- B1, EfficientNetV2-B2, EfficientNetV2-B3, EfficientNetV2-S and EfficientNetV2-M were stacked.

EROL et al. [23] made a comparison between three types of ensemble techniques; namely, bagging, AdaBoost and random forest, to detect COVID-19. Adaboost classifier achieved the highest accuracy of 97.25%. Bagging and random-forest classifiers achieved an accuracy of 96.69% and 96.89%, respectively. The BIGDATA-COVID19 dataset was used that includes age, sex and routine blood-test results of 1218 patients.

Banerjee et al. [24] presented the blending ensemble technique of DenseNet-201 snapshots, providing a variety of information regarding the features that were taken out of CXRs to detect COVID-19. To merge the decision scores, they employed the decision-level fusion approach, which involves a Random Forest (RF) meta-learner and the blending method. On the large COVID-X dataset, the model achieved an accuracy score of 94.55% and on the smaller dataset by Chowdhury et al., the model achieved an accuracy score of 98.13%.

There are some common difficulties/issues in the papers that we presented, because of which the reliability of the proposed deep-learning models can be questioned as data imbalance, image-size handling, dataset availability and high correlation of errors when employing ensemble techniques. The potential ways to overcome these issues include further experimentation, data augmentation, ensemble-model refinement and feature engineering to extract more relevant features from the CXR images.

3. PROPOSED WORK

3.1 Dataset

The COVID-19 radiography database of chest X-ray images is one of the available datasets utilized to develop and evaluate deep-learning models for the detection and classification of lung diseases, particularly COVID-19, as shown in Figure 1. The database consists of a collection of images from multiple sources, such as the COVID-19 Image Dataset, the COVID-19 Database of the Italian Society of Medical and Interventional Radiology (SIRM) and images from several different publications [25]-[26]. It includes 3616 COVID-19-positive cases, 10,192 Normal, 6012 Lung Opacity (Non-COVID lung infection) and 1345 Viral Pneumonia images in PNG format [27]. All images have the same resolution of 299×299. The dataset was split into training, validation and test sets. The training set was used to train the model, the validation set was used to validate and tune the hyper-parameters of the model and the test set was used to evaluate the overall performance of the model. The training set

included 1075 normal images, 1075 images of COVID-19 and 1075 images of pneumonia. The validation set and test set included 135 normal images, 135 COVID-19 images and 135 pneumonia images.

The COVIDx CXR-4 dataset is an open-source benchmark dataset that combines 5 different publicly available datasets which include: COVID-19 Image data collection, COVID-19 Chest X-Ray Dataset Initiative, COVID-19 radiography database, RSNA Pneumonia Detection challenge dataset and ActualMed COVID-19 Chest X-Ray Dataset Initiative [28]. The dataset includes 84,818 images from 45,342 patients in PNG format with the same resolution of 1024×1024. The dataset is classified into positive and negative COVID-19 samples. We split the data into training, validation and test sets. The training set includes 9600 images, the validation set includes 1200 images and the test set includes 1200 images. Each set contains 3 classes: COVID-19, normal and pneumonia.

Figure 1. A batch of images from the training dataset.

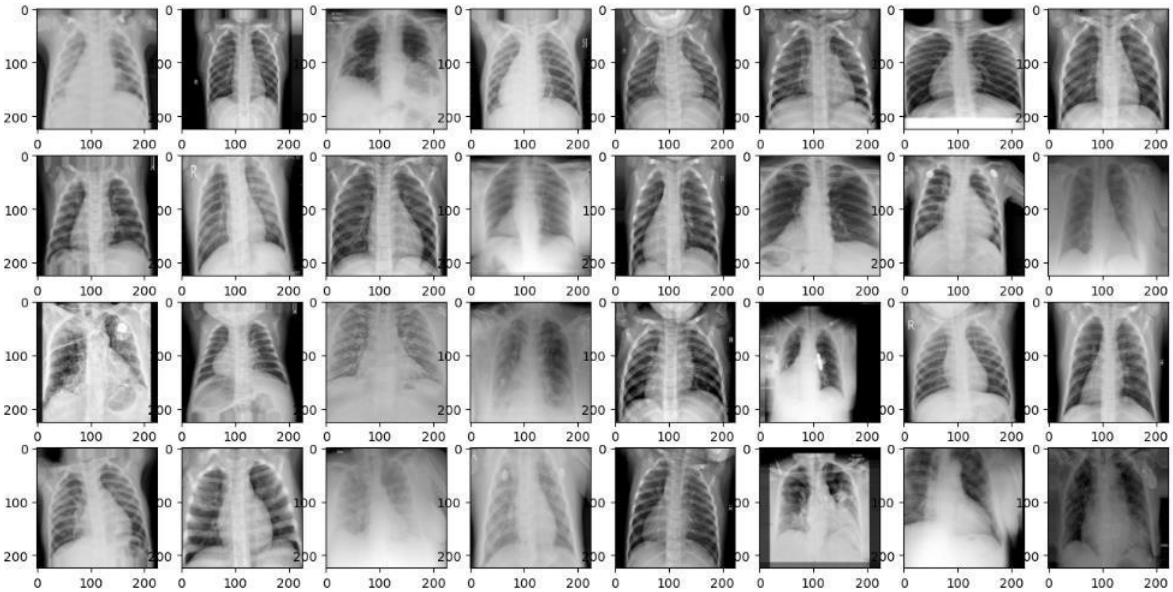


Figure 2. A batch of training images after data augmentation.

3.2 Data Preparation

A data-augmentation technique is used to increase data diversity, improve generalization and reduce overfitting in deep-learning models. Random transformations, such as rescale, shear, zoom and horizontal flip, were applied to the training images as shown in Figure 2. Both the shear-range and zoom-range parameters were set to 0.2 in the ImageDataGenerator class from the Keras library. Images have been resized to dimensions of 224×224.

3.3 Model

This research is a continuation of previous work where we used both DenseNet-169 and vision transformer (ViT-l32) models to detect COVID-19 and pneumonia lung diseases. The results were as follows: the accuracy of both models was 92.31% and 92.56%, respectively. To improve the

performance of the two models, they were combined using two different types of ensemble techniques: decision-level fusion and feature-level fusion. Indeed, the accuracy improved to 93.3% and 94.54%, respectively [3]. Now we use other ensemble methods on the same two models and the same dataset to achieve a better-performance model.

3.3.1 DenseNet-169

A CNN architecture called DenseNet-169 was created specifically for image-classification applications. It belongs to the DenseNet model family, which is famous for having dense connectivity between layers. It is composed of several dense blocks as shown in Figure 3, each dense block consisting of several convolutional layers. The primary advantage of DenseNet is its dense connection architecture, where each layer is feed-forward connected to every other layer. Enhanced information flow and feature reuse across the network are made possible by this dense connectivity. To reduce the number of parameters in the network and reduce the spatial dimensions of feature maps, transition layers between dense blocks are used. Transition layers include a combination of convolutional layers, pooling layers, batch-normalization and nonlinear-activation functions. When compared to the DenseNet-121 model, the DenseNet-169 is larger and more accurate. It is about 55MB in size and contains about 169 layers [29]. The increased depth allows DenseNet-169 to capture more complex features and potentially achieve improved accuracy in image-classification tasks [30]. The DenseNet-169 model takes an image with a size of 224×224 pixels as input [31]. Compared with other CNN architectures, it is relatively low in parameters.

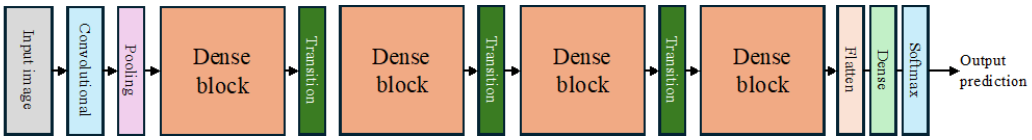


Figure 3. DenseNet-169 architecture.

3.3.2 Vision Transformer

The vision transformer (ViT) has shown promising effectiveness for global feature extraction in many tasks of computer vision, such as image recognition, image classification, object detection and image segmentation [32]. It has gained attention due to its ability to capture long-term dependencies and its generalizability across different data modalities. The ViT model is an effective tool for image classification, because it uses self-attention processes to obtain global information from an image. By capturing global and local representations from shallow and deep layers, the ViT model differs from traditional CNN, which concentrates mostly on local features using convolutional filters [33], but ViT processes images using patches and the self-attention method by converting input images into a sequence of tokens. The input image is split up into fixed-size patches in the first stage. Every patch is viewed as a token and is put through a linear embedding process. Subsequently, position embeddings are appended to the patch embeddings to furnish spatial details regarding the patches' placement within the image. The sequences of patch embeddings and position embeddings are then put into a typical transformer encoder. The transformer encoder is composed of feed-forward neural networks and numerous layers of self-attention. By creating attention maps from the provided embedded visual tokens, the multi-head self-attention (MSA) enables the model to focus on several regions of the input image concurrently. Batch normalization is used to improve training stability and reduce training time. The residual connections improve the overall performance of the network [34]. Figure 4 illustrates the ViT architecture.

There are several variations related to ViT as ViT-l16, ViT-l32, ViT-b16, ViT-b32 and data-efficient image transformers (DeiT). The ViT-l32 is considered more significant and more powerful than some of the other variants [35]-[36] due to its ability to achieve superior results in image-recognition tasks, so ViT-l32 was chosen. l means 'large' and 32 refers to batch size.

3.3.3 The Used Ensemble Techniques Background

The ensemble techniques refer to the use of several base models and combining their predictions to improve the overall performance and accuracy of the system [37]. Instead of relying on a single model, ensemble techniques leverage the diversity and collective intelligence of multiple models to

produce one optimal predictive model. It can reduce bias, variance and overfitting [38]-[39]. There are several types of ensemble techniques as bagging, stacking, blending and boosting [40].

In this paper, we applied a comprehensive range of ensemble techniques, including Bagging, Gradient Boosting, AdaBoost, XGBoost, LightGBM, CatBoost, Stacking and Blending.

Bagging, also known as bootstrap aggregation, is an ensemble method that involves training multiple models independently on random sub-sets of the data [41]-[43]. The bagging involves the following steps: (1) Generating predictions from the base models using the holdout set (test set). (2) Combining the predictions as features for the bagging model using concatenate function. (3) Training a bagging model using the combined features and true labels. (4) Generating predictions from the trained bagging model on the combined features. (5) Evaluating the bagged model performance using the accuracy_score function with the combined predictions.

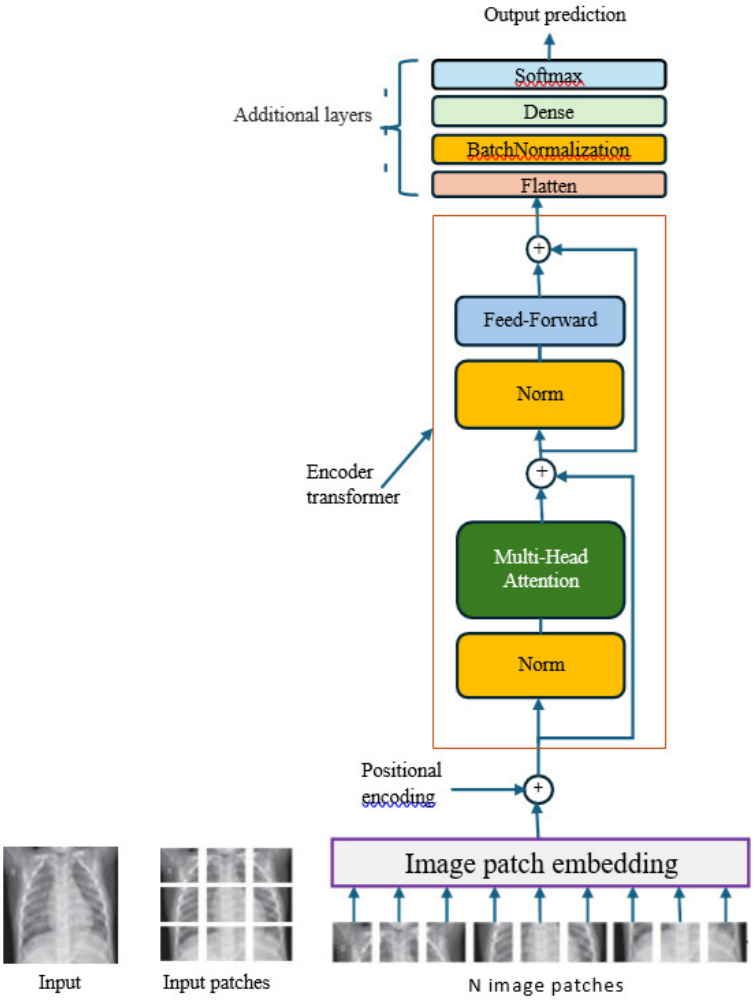


Figure 4. ViT-l32 structure.

Stacking uses a meta-learning algorithm to learn how to best combine the predictions from two or more models [44]. It can combine the capabilities of a group of models that perform well on a classification or regression task and make predictions that perform better than any single model in the group [45]-[46]. The stacking involves the following steps: (1) Compiling the models and making predictions. (2) Combining the predictions from the base models to create meta-features. (3) Useing the combined meta-features as input features for the meta-learner. (4) Training the meta-learner model using the meta-features and the true labels. (5) Evaluating the performance of the stacked model using appropriate evaluation metrics. This implementation follows the stacking ensemble technique, where predictions from base models are combined using a meta-model to improve predictive performance. The use of a meta-model allows for the aggregation of predictions from diverse base models, aiming to reduce overfitting and improve generalization.

Boosting is designed to address potential limitations of individual models, such as generalizability and data bias. By combining the predictions of multiple models, ensemble methods aim to improve the robustness and generalization of machine-learning models by combining a group of weak learners into a strong learner. It trains the models sequentially, where each subsequent model focuses on correcting the mistakes made by the previous models, so it can minimize training errors [47]-[48]. There are many types of boosting algorithms, such as AdaBoost, CatBoost, XGBoost and LGBoost [49]-[50].

The short form for adaptive boosting (AdaBoost) is a powerful ensemble learning algorithm used in machine learning for addressing binary-classification challenges.

The catBoost is a high-performance open-source library for gradient boosting on decision trees, designed for use in classification, regression and ranking tasks.

The eXtreme Gradient Boosting (XGBoost) is an open-source software library that provides a regularizing gradient boosting framework for various programming languages, such as C++, Java and Python.

The light gradient boosting machine (LGBoost) is an open-source gradient boosting framework that is designed for efficiency, scalability and high performance in machine-learning tasks.

The boosting steps: (1) Training each model separately. (2) Generating predictions from each model. (3) Combining the predictions using the concatenate function. (4) Training a boosting model (GradientBoost- ingClassifier) using the stacked features. (5) Generating predictions from the trained boosting model on the stacked features. (6) Evaluate the performance of the boosted model using appropriate evaluation metrics.

Blending is an ensemble approach that can improve the model performance to be more accurate. It uses a particular method to merge predictions from various models contributing to the ensemble. The steps used in the blending process are: (1) Generating the predictions from the base models. (2) Building the model from the test set and the predictions. (3) The building model serves as the meta-model. (4) Generating predictions from the meta-model.

4. RESULTS AND DISCUSSION

The proposed method was implemented using Python 3.8 with additional libraries, such as Pandas, Tensor Flow and Keras. Windows 10 Operating System powered the System with configuration, Intel(R) Xeon(R) CPU E5-2687W v3 @ 3.10 GHz, NVIDIA GeForce GTX 970 GPU and 64 GB RAM.

The first dataset "COVID-19 Radiography dataset" obtained from Kaggle [27], was utilized to train and test DenseNet-169 and ViT-l32 models for multi-level classification aimed at detecting COVID-19 patients. The training and validation sets comprised 90% of the dataset, while the testing set utilized the remaining 10%, as outlined in Table 1. Python and machine-learning libraries were employed for implementation, with the Python programming language utilized to train and evaluate the proposed models, which were pre-trained using TensorFlow. The training data underwent modification through data-augmentation techniques, as illustrated in Figure 2. The second dataset COVIDx CXR-4 was split into the training, validation and test set in the ratio 8:1:1, as shown in Table 2.

The pre-trained DenseNet 169 model was trained on the initialization weights illustrated in Table 3 using the first dataset and the Adam optimizer. Subsequently, the ViT-l32 was separately trained on the same dataset. Predictions were generated from the two models using the test set and combined using ensemble techniques, as shown in Figure 5.

Table 1. Class-wise distribution of CXR samples in the COVID-19 Radiography database.

Phase	COVID-19	Normal	Pneumonia	Total
Training	1075	1075	1075	3225
Validation	135	135	135	405
Test	135	135	135	405
Total	1345	1345	1345	4035

Table 2. Class-wise distribution of CXR samples in the COVIDx CXR-4 dataset.

Phase	COVID-19	Normal	Pneumonia	Total
Training	3200	3200	3200	9600
Validation	400	400	400	1200
Test	400	400	400	1200
Total	4000	4000	4000	12000

Table 3. Training parameters.

Training Parameters	**Values/Types**
Number of epochs	100
Batch size	32
Optimizer	Adam
Learning rate	0.001
Zoom and shear range	20%
Fill mode	Nearest
Rescale	1./255
Horizontal flip	True
Shuffle	True
Class mode	Categorical

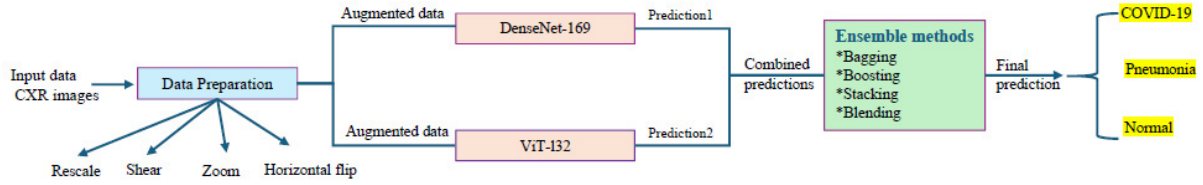


Figure 5. Proposed workflow model.

Ensemble techniques typically refer to Bagging (bootstrap-aggregating), Boosting or Stacking/Blending techniques to induce high variability among the base models. These techniques aim to combine the predictions from multiple models to improve predictive performance. To evaluate the bagged-model performance, we combined the predictions from the two models using a voting mechanism (majority voting). Finally, we evaluated the combined predictions and obtained an accuracy of 98.27% from the random-forest bagging model, as shown in Figure 6a.

When implementing boosting ensemble techniques using the DenseNet169 and ViTl32 models, the weak learners (base models) are combined sequentially to form a strong learner (ensemble model). The gradient- boost, AdaBoost, XGBoost, LGBoost and CatBoost ensemble techniques achieved an accuracy of 98.765%, 97.04%, 96.54%, 97.79% and 99.753%, respectively as shown in Figure 6b, Figure 6c, Figure 6d, Figure 6e and Figure 6f.

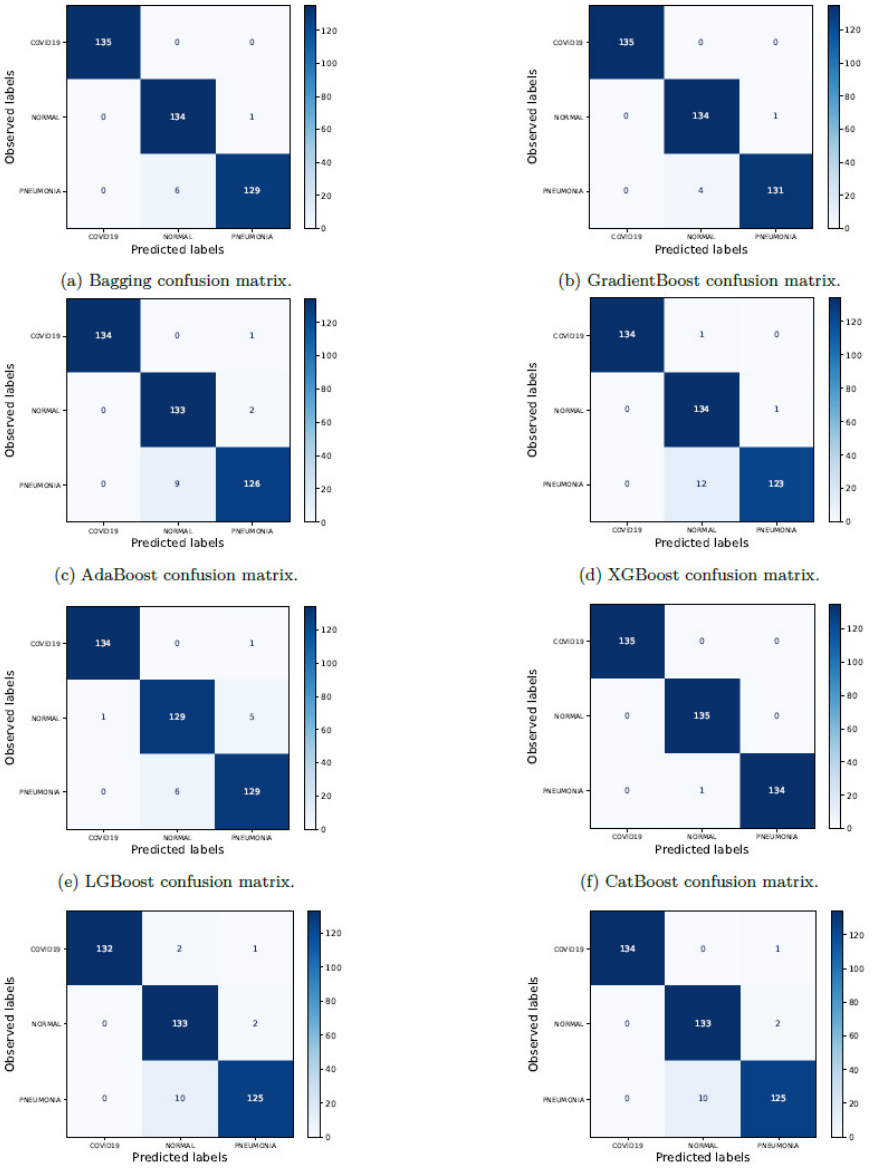
Stacking, also known as stacked generalization, allows a training algorithm to ensemble several similar learning-algorithm predictions. A stacking model is implemented using a holdout set to generate predictions from base models (CNN and ViTl32 models). These predictions are then concatenated to create a stacked dataset. The true labels for the holdout set are one-hot encoded using the OneHotEncoder class. A meta-model for multi-class classification is defined and trained using the stacked dataset. The meta-model consists of three dense layers with ReLU and Softmax activations. The accuracy and F1-score of the stacked ensemble model are then calculated. The accuracy result was 96.296%, as shown in Figure 6g.

The blending ensemble technique combines the predictions of several base models to enhance overall predictive performance, minimize overfitting and leverage the advantages of different methods. It achieved an accuracy of 96.79%, as shown in Figure 6h.

So, from the previous results, CatBoost achieved the highest accuracy using one dataset. The optimization for the ensemble implementation involved: 1) A combination of hyper-parameter tuning,

such as the learning rate, depth and regularization parameters. 2) Model selection. 3) Validation and benchmarking against other algorithms to achieve the highest accuracy for the specific dataset.

Other performance metrics, such as precision, recall, f1-score, sensitivity, specificity and ROC-AUC score, can be calculated using the COVID-19 Radiography dataset and the results are shown in Table 4.



(a) Bagging confusion matrix. (b) GradientBoost confusion matrix.

(c) AdaBoost confusion matrix. (d) XGBoost confusion matrix.

(e) LGBoost confusion matrix. (f) CatBoost confusion matrix.

(g) Stacking confusion matrix. (h) Blending confusion matrix.

Figure 6. Confusion matrices for all ensemble models using COVID-19 Radiography dataset.

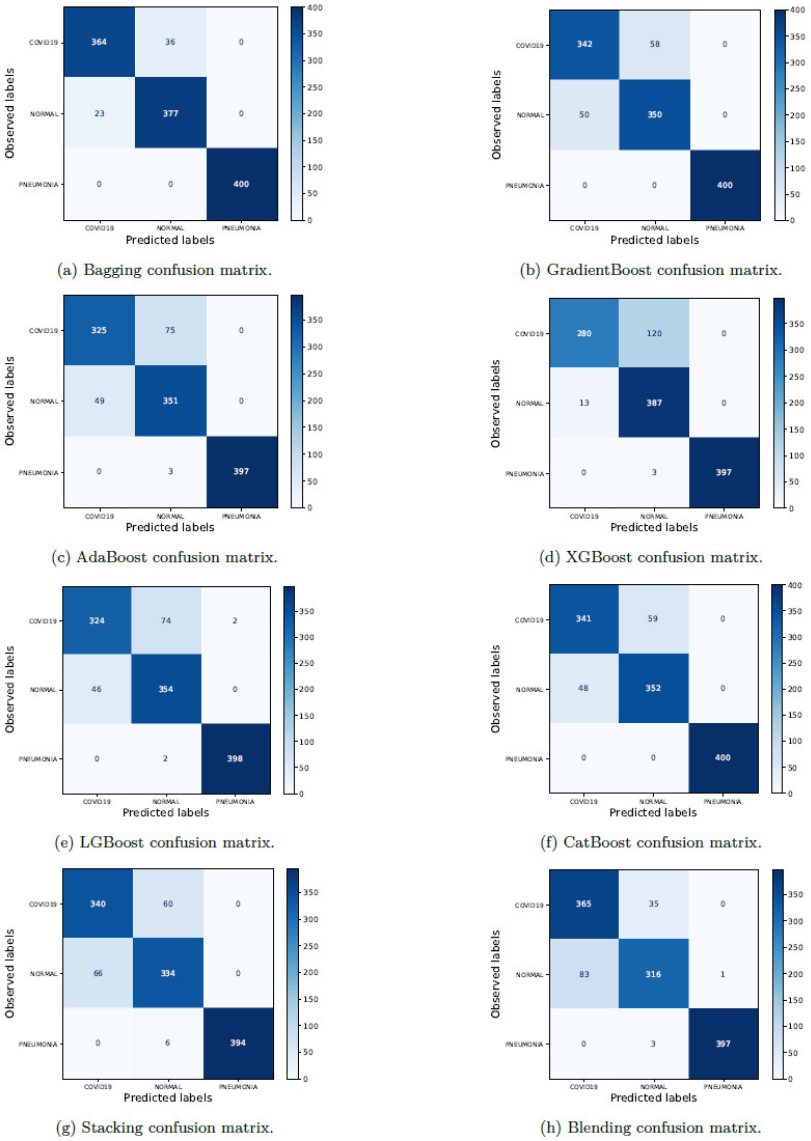
Table 4. Performance of the proposed ensemble models using the COVID-19 Radiography dataset.

Ensemble Techniques	Precision	Recall	F1-score	Sensitivity	Specificity	ROC-AUC score
Bagging	98.53%	98.52%	98.52%	100%	100%	99.69%
GradientBoost	98.55%	98.52%	98.52%	100%	100%	99.94%
AdaBoost	97.79%	97.78%	97.78%	100%	100%	99.54%
XGBoost	97%	96.79%	96.80%	100%	99.26%	99%
LGBoost	97.55%	97.53%	97.53%	100%	100%	99.84%
CatBoost	99.51%	99.51%	99.51%	100%	100%	99.99%
Stacking	99.25%	100%	96.06%	100%	99.25%	99.69%
Blending	99.25%	100%	96.06%	100%	99.25%	99.72%

When using more than one dataset (COVIDx CXR-4) the results became as follows: RF bagging model achieved an accuracy of 95.08%, as shown in Figure 7a. The gradient-boost, AdaBoost, XGBoost, LGBoost and CatBoost ensemble techniques achieved an accuracy of 91%, 89.42%, 88.67%, 89.67% and 91.08%, respectively, as shown in Figure 7b, Figure 7c, Figure 7d, Figure 7e and Figure 7f. The stacking and blending models achieved an accuracy of 89% and 89.83%, as shown in Figure 7g and Figure 7h, respectively. From these results, the RF bagging ensemble model achieved the highest accuracy of 95.42%. The other performance metrics using COVIDx CXR-4 can be calculated and the results are shown in Table 5.

Table 5. Performance of the proposed ensemble models using COVIDx CXR-4.

Ensemble Techniques	Precision	Recall	F1-score	Sensitivity	Specificity	ROC-AUC score
Bagging	95.17%	95.17%	95.2%	92%	93.5%	99.69%
GradientBoost	92.45%	92.42%	92.41%	86.75%	90.5%	98.6%
AdaBoost	91.2%	91.1%	91.2%	83.5%	90.68%	96.37%
XGBoost	91.15%	91.17%	90.1%	95.25%	75.25%	96.45%
LGBoost	91.67%	91.67%	91.65%	86.25%	89.89%	98.53%
CatBoost	93.1%	93.1%	93.1%	89%	90.25%	98.79%
Stacking	91.94%	91.92%	91.92%	86.5%	89.5%	98.35%
Blending	91.3%	91.25%	91.25%	89%	85%	98.34%



(a) Bagging confusion matrix. (b) GradientBoost confusion matrix.

(c) AdaBoost confusion matrix. (d) XGBoost confusion matrix.

(e) LGBoost confusion matrix. (f) CatBoost confusion matrix.

(g) Stacking confusion matrix. (h) Blending confusion matrix.

Figure 7. Confusion matrices for all ensemble models using COVIDx CXR-4 dataset.

Table 6. Comparison between the proposed model and previous ensemble studies.

Model	Dataset	Number of images	Accuracy
Weighted average ensemble model [16]	Chest X-ray dataset	The dataset is divided into 2 classes: Normal and Pneumonia. 4192 for the training set, 1040 for the validation set, and 624 for the test set.	92%
Weighted average ensemble model [17]	COVIDx dataset	The dataset is divided into 3 classes: Normal, Pneumonia, and COVID-19. 13898 for the training set and 1579 for the test set.	95%
XGBoost classifier [18]	Chest X-ray image database	The dataset is divided into 2 classes: Multi-Drug Resistant Tuberculosis (MDR–TB) and Drug Sensitive Tuberculosis (DS-TB). 199 DS-TB and 202 MDR-TB.	93%
Boosted model [19]	New dataset collected from different sources	The dataset is divided into 3 classes: COVID-19, Pneumonia, and Normal. 6755 for the training set, 1822 for the validation set, and 911 for the test set.	99.35%
Stacked model [20]	COVID-19 image data collection repository and chest X-ray images	The dataset is divided into 2 classes: Normal and Pneumonia. COVID-19 image data collection repository: 286 images for training and 192 for testing. chest X-ray images dataset: 3900 images for training and 624 for testing.	94.24%
Stacked model [22]	COVID-19 Radiography Dataset, Chest X-Ray Images (Pneumonia), and Tuberculosis Chest X-ray Dataset	The dataset is divided into 5 classes: COVID-19, Bacterial pneumonia, Tuberculosis, Viral pneumonia, and Normal. In training, validation, and test subsets: 900, 100, and 200, respectively.	99.21%
Adaboost classifier [23]	BIGDATA-COVID19	The dataset is divided into 2 classes: Severity and Dead. 3996 for the training set and 999 for the test set.	97.25%
Blending model [24]	(COVID-X and Chowdhury datasets	The dataset is divided into 3 classes: Normal, Pneumonia, and COVID-19. The COVID-X dataset: 11115 for the training set, 1579 for the validation set, and 2777 for the test set. The COVID dataset by Chowdhury et al.: 1807 for the training set, 321 for the validation set, and 777 for the test set.	98.13%
Decision-level fusion model [3]	COVID-19 Radiography database	The dataset is divided into 3 classes: Normal, Pneumonia, and COVID-19. 3225 images for the training set, 405 for the validation set, and 403 for the test set.	93.3%
Feature-level fusion model [3]	COVID-19 Radiography database	The dataset is divided into 3 classes: Normal, Pneumonia, and COVID-19. 3225 images for the training set, 405 for the validation set, and 403 for the test set.	94.54%
Proposed RF bagging model	COVID-19 Radiography dataset	3225 images for the training set, 405 for the validation set, and 405 for the test set.	98.27%
Proposed GradientBoost model	COVID-19 Radiography dataset	3225 images for the training set, 405 for the validation set, and 405 for the test set.	98.765%
Proposed AdaBoost model	COVID-19 Radiography dataset	3225 images for the training set, 405 for the validation set, and 405 for the test set.	97.04%
Proposed XGBoost model	COVID-19 Radiography dataset	3225 images for the training set, 405 for the validation set, and 405 for the test set.	96.54%
Proposed LGBoost model	COVID-19 Radiography dataset	3225 images for the training set, 405 for the validation set, and 405 for the test set.	96.79%
Proposed CatBoost model	COVID-19 Radiography dataset	3225 images for the training set, 405 for the validation set, and 405 for the test set.	99.753%
Proposed stacking model	COVID-19 Radiography dataset	3225 images for the training set, 405 for the validation set, and 405 for the test set.	96.3%
Proposed blending model	COVID-19 Radiography dataset	3225 images for the training set, 405 for the validation set, and 405 for the test set.	96.79%
Proposed RF bagging model	COVIDx CXR-4 dataset	9600 images for the training set, 1200 for the validation set, and 1200 for the test set.	95.08%
Proposed GradientBoost model	COVIDx CXR-4 dataset	9600 images for the training set, 1200 for the validation set, and 1200 for the test set.	91%
Proposed AdaBoost model	COVIDx CXR-4 dataset	9600 images for the training set, 1200 for the validation set, and 1200 for the test set.	89.42%
Proposed XGBoost model	COVIDx CXR-4 dataset	9600 images for the training set, 1200 for the validation set, and 1200 for the test set.	88.67%
Proposed LGBoost model	COVIDx CXR-4 dataset	9600 images for the training set, 1200 for the validation set, and 1200 for the test set.	89.67%
Proposed CatBoost model	COVIDx CXR-4 dataset	9600 images for the training set, 1200 for the validation set, and 1200 for the test set.	91.08%
Proposed stacking model	COVIDx CXR-4 dataset	9600 images for the training set, 1200 for the validation set, and 1200 for the test set.	89%
Proposed blending model	COVIDx CXR-4 dataset	9600 images for the training set, 1200 for the validation set, and 1200 for the test set.	89.83%

By looking at the results summarized in Figure 8, we find that the accuracy of the models decreased when using more than one dataset, which may be due to various factors, such as data complexity and diversity. As previously mentioned, this work is a continuation of our last work [3] to improve the model's performance. The results we obtained from the earlier work were as follows: decision-level fusion and feature-level fusion achieved an accuracy of 93.3% and 94.53%, respectively. However, the results from the advanced ensemble techniques reached 99.753% when using the same COVID-19 Radiograph dataset. Finally, the results confirm that the performance of these advanced ensemble models surpasses that of fusion models, as shown in Figure 8. Table 6 compares the previous studies and our proposed methods.

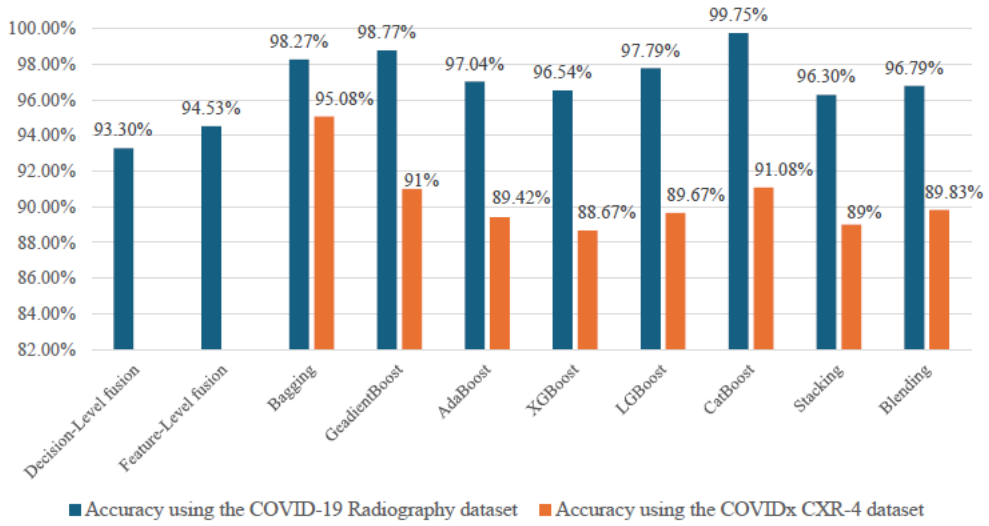


Figure 8. The accuracy of our proposed ensemble models.

5. Conclusions and Future Research Directions

This paper presented different types of advanced ensemble techniques to improve the model performance in diagnosing lung diseases using CXR images. We used pre-trained CNN models DenseNet-169 and ViT-l32. This work is related to a previous research paper, but we presented different methods that showed more accurate results. In previous work, we used simple ensemble techniques, such as feature-level fusion and decision-level fusion, achieving accuracy results as follows: 94.54% and 93.3%, respectively. However, using advanced ensemble techniques, we achieved a higher accuracy of fusion operations, reaching 99.753%. To aid in lung-disease prevention and early diagnosis, researchers continue to develop a variety of detection technologies and architectures by increasing the size and diversity of training datasets, but this can be costly and time-consuming. To avoid these issues, researchers are exploring techniques like data augmentation to address the challenge of limited datasets.

In our future work, we plan to use multiple datasets and explainable AI (XAI) models to enhance the accuracy and comprehensiveness of lung-disease diagnosis and classification. We also consider including multi-model data to expand the feature space and improve disease-classification accuracy, such as medical records and clinical metadata.

References

[1] F. A. Mostafa, L. A. Elrefaei, M. M. Fouda and A. Hossam, "A Survey on AI Techniques for Thoracic Diseases Diagnosis Using Medical Images," Diagnostics, vol. 12, no. 12, p. 3034, 2022.

[2] I. Team, "Coronavirus Cases," [Online]. Available: https://www.worldometers.info/ coronavirus/, 2024.

[3] F. A. Mostafa, L. A. Elrefaei, M. M. Fouda and A. Hossam, "Diagnosis of Lung Diseases from Chest X- Ray Images Using Different Fusion Techniques," Proc. of the 2023 11th Int. Conf. on Information and Communication Technology (ICoICT), pp. 429–435, Melaka, Malaysia, 2023.

[4] M. H. Saad, S. Hashima, W. Sayed, E. H. El-Shazly, A. H. Madian and M. M. Fouda, "Early Diagnosis of COVID-19 Images Using Optimal CNN Hyper-parameters," Diagnostics, vol. 13, no. 1, p. 76, 2023.

[5] Z. Ahmad, S. Rahim, M. Zubair and J. Abdul-Ghafar, "Artificial Intelligence (AI) in Medicine, Current Applications and Future Role with Special Emphasis on Its Potential and Promise in Pathology: Present and Future Impact, Obstacles Including Costs and Acceptance among Pathologists, Practical and Philosophical Considerations: A Comprehensive Review," Diagnostic pathology, vol. 16, pp. 1–16, 2021.

[6] H. Q. Nguyen et al., "VinDr-CXR: An Open Dataset of Chest X-Rays with Radiologist's Annotations," Scientific Data, vol. 9, no. 1, p. 429, 2022.

[7] A. Shrestha and A. Mahmood, "Review of Deep Learning Algorithms and Architectures," IEEE Access, vol. 7, pp. 53040–53065, 2019.

[8] Z. Yu, K. Wang, Z. Wan, S. Xie and Z. Lv, "Popular Deep Learning Algorithms for Disease Prediction: A Review," Cluster Computing, vol. 26, no. 2, pp. 1231–1251, 2023.

[9] A. A. Chowdhury, K. T. Hasan and K. K. S. Hoque, "Analysis and Prediction of COVID-19 Pandemic in Bangladesh by Using ANFIS and LSTM Network," Cognitive Computation, vol. 13, pp. 761-770, [Online], Available: https://link.springer.com/article/10.1007/s12559-021-09859-0, Apr. 2021.

[10] D. Kuzinkovas and S. Clement, "The Detection of COVID-19 in Chest X-Rays Using Ensemble CNN Techniques," Information, vol. 14, no. 7, p. 370, 2023.

[11] A. Zizaan and A. Idri, "Evaluating and Comparing Bagging and Boosting of Hybrid Learning for Breast Cancer Screening," Scientific African, vol. 23, p. e01989, DOI: 10.1016/j.sciaf.2023.e01989, 2024.

[12] U. Ahmed, J. C.-W. Lin and G. Srivastava, "Towards Early Diagnosis and Intervention: An Ensemble Voting Model for Precise Vital Sign Prediction in Respiratory Disease," IEEE Journal of Biomedical and Health Informatics, pp. 1-13, DOI: 10.1109/JBHI.2023.3270888, 2023.

[13] U. Bhimavarapu, N. Chintalapudi and G. Battineni, "Multi-classification of Lung Infections Using Improved Stacking Convolution Neural Network," Technologies, vol. 11, no. 5, p. 128, 2023.

[14] T. Mahesh et al., "Blended Ensemble Learning Prediction Model for Strengthening Diagnosis and Treatment of Chronic Diabetes Disease," Computational Intelligence and Neuroscience, vol. 2022, DOI: 10.1155/2022/4451792, 2022.

[15] T. Mahesh et al., "Early Predictive Model for Breast Cancer Classification Using Blended Ensemble Learning," Int. J. of System Assurance Eng. and Management, vol. 15, no. 1, pp. 188–197, 2024.

[16] R. Hasan, S. M. Azmat Ullah, A. Nandi and A. Taher, "Improving Pneumonia Diagnosis: A Deep Transfer Learning CNN Ensemble Approach for Accurate Chest X-ray Image Analysis," Proc. of the 2023 Int. Conf. on Information and Communication Technology for Sustainable Development (ICICT4SD), pp. 109–113, hal-04163800f, 2023.

[17] S. Tang, C. Wang, J. Nie, N. Kumar, Y. Zhang, Z. Xiong and A. Barnawi, "EDL-COVID: Ensemble Deep Learning for COVID-19 Case Detection from Chest X-ray Images," IEEE Transactions on Industrial Informatics, vol. 17, no. 9, pp. 6539–6549, 2021.

[18] S. Govindarajan, S. R. Manuskandan and R. Swaminathan, "Diagnostics of Multi Drug Resistant Tuberculosis in Chest Radiographs Using Local Textures & Extreme Gradient Boosting," Current Directions in Biomedical Engineering, vol. 9, no. 1, pp. 721–724, 2023.

[19] S. Kalaivani and K. Seetharaman, "A Three-stage Ensemble Boosted Convolutional Neural Network for Classification and Analysis of COVID-19 Chest X-ray Images," Int. Journal of Cognitive Computing in Engineering, vol. 3, pp. 35–45, 2022.

[20] R. Soundrapandiyan, H. Naidu, M. Karuppiah, M. Maheswari and R. C. Poonia, "AI-based Wavelet and Stacked Deep Learning Architecture for Detecting Coronavirus (COVID-19) from Chest X-ray Images," Computers and Electrical Engineering, vol. 108, Article no. 108711, 2023.

[21] P. Mooney, "Chest X-ray Images (Pneumonia)," Kaggle, [Online], Available: https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia, Mar. 2018.

[22] M.-L. Huang and Y.-C. Liao, "Stacking Ensemble and ECA-EfficientNetV2 Convolutional Neural Networks on Classification of Multiple Chest Diseases Including COVID-19," Academic Radiology, vol. 30, no. 9, pp. 1915–1935, 2023.

[23] G. Erol and B. Uzbaş, "Detection of COVID-19 Severity and Mortality from Blood Parameters by Ensemble Learning Methods," Karaelmas Fen ve Mühendislik Dergisi, vol. 13, p. 316–328, 2023.

[24] A. Banerjee, A. Sarkar, S. Roy, P. K. Singh and R. Sarkar, "Covid-19 Chest X-ray Detection through Blending Ensemble of CNN Snapshots," Biomedical Signal Processing and Control, vol. 78, p. 104000, 2022.

[25] Peshotan, "Summary of COVID-19 Radiography Database," [Online], Available: Https://github.com/sfu-db/covid19-datasets/blob/master/datasets-details/radiography-kaggle-COVID-19-dataset.md, 2020.

[26] T. Pham, "Classification of COVID-19 Chest X-rays with Deep Learning: New Models or Fine Tuning?" Health Information Science and Systems, vol. 9, no. 1, DOI: 10.1007/s13755-020-00135-3, 2020.

[27] T. Rahman, "Covid-19 Radiography Database," [Online], Available: https: /www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database, Mar. 2022.

[28] Y. Wu, H. Gunraj, C. en Amy Tai and A. Wong, "COVIDx CXR-4: An Expanded Multi-institutional Open-source Benchmark Dataset for Chest X-ray Image-based Computer-aided COVID-19 Diagnostics," arViv: 2311.17677, DOI: arxiv-2311.17677, 2023.

[29] P. P. Dalvi, D. R. Edla and B. R. Purushothama, "Diagnosis of Coronavirus Disease from Chest X-ray Images Using Densenet-169 Architecture," SN Computer Science, vol. 4, Article no. 214, [Online], Available: https://link.springer.com/article/10.1007/s42979-022-01627-7, Feb. 2023.

[30] M. K. Bohmrah and H. Kaur, "Classification of COVID-19 Patients Using Efficient Fine-tuned Deep Learning DenseNet Model," Global Transitions Proceedings, vol. 2, no. 2, pp. 476–483, 2021.

[31] A. Ala and Polat, "Detection of COVID-19 from Computed Tomography Images with DenseNet based Deep Learning Models," Proc. of the 2021 29th Signal Processing and Communications Applications Conference (SIU), pp. 1-4, Istanbul, Turkey, 2021.

[32] K. Han et al., "A Survey on Vision Transformer," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 1, pp. 87–110, 2023.

[33] S. Park et al., "Multi-task Vision Transformer Using Low-level Chest X-ray Feature Corpus for COVID-19 Diagnosis and Severity Quantification," Medical Image Analysis, vol. 75, Article no. 102299, 2022.

[34] C. C. Ukwuoma et al., "Automated Lung-related Pneumonia and COVID-19 Detection Based on Novel Feature Extraction Framework and Vision Transformer Approaches Using Chest X-ray Images," Bioengineering, vol. 9, no. 11, Article no. 709, 2022.

[35] T. Aitazaz, A. Tubaishat, F. Al-Obeidat, B. Shah, T. Zia and A. Tariq, "Transfer Learning for Histopathology Images: An Empirical Study," Neural Computing and Applications, vol. 35, pp. 7963-7974, Jul. 2022.

[36] G. Boesch, "Vision Transformers (ViT) in Image Recognition," 2024 Guide, [Online], Available: https://viso.ai/deep-learning/vision-transformer-vit/, Jun. 2024.

[37] R. Kundu, R. Das, Z. W. Geem, G.-T. Han and R. Sarkar, "Pneumonia Detection in Chest X-ray Images Using an Ensemble of Deep Learning Models," PloS One, vol. 16, no. 9, p. e0256630, 2021.

[38] S. Kaleem, A. Sohail, M. U. Tariq, M. Babar and B. Qureshi, "Ensemble Learning for Multi-class COVID-19 Detection from Big Data," Plos One, vol. 18, no. 10, p. e0292587, 2023.

[39] O. O. Abayomi-Alli et al., "An Ensemble Learning Model for COVID-19 Detection from Blood Test Samples," Sensors, vol. 22, no. 6, Article no. 2224, 2022.

[40] S. Shukla, K. Arya, B. Garg and P. Lezama, "Pneumonia Detection Using Gradient Boosting on Selected Features Extracted from DenseNet," Proc. of the 7th ASRES Int. Conf. on Intelligent Technologies, in Book Series: Lecture Notes in Networks and Systems, vol. 685, pp. 181–195, 2022.

[41] A. S. Ashour, M. M. Eissa, M. A. Wahba, R. A. Elsawy, H. F. Elgnainy, M. S. Tolba and W. S. Mohamed, "Ensemble-based Bag of Features for Automated Classification of Normal and COVID-19 CXR Images," Biomedical Signal Processing and Control, vol. 68, p. 102656, 2021.

[42] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince and F. Herrera, "A Review on Ensembles for the Class Imbalance Problem: Bagging-, Boosting- and Hybrid-based Approaches," IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews), vol. 42, no. 4, pp. 463–484, 2011.

[43] R. Odegua, "An Empirical Study of Ensemble Techniques (Bagging, Boosting and Stacking)," Proc. of Deep Learning Conf. (IndabaXAt), DOI: 10.13140/RG.2.2.35180.10882, 2019.

[44] Z. Hu, H. Qiu, Z. Su, M. Shen and Z. Chen, "A Stacking Ensemble Model to Predict Daily Number of Hospital Admissions for Cardiovascular Diseases," IEEE Access, vol. 8, pp. 138719–138729, 2020.

[45] M. Gour and S. Jain, "Automated COVID-19 Detection from X-ray and CT Images with Stacked Ensemble Convolutional Neural Network," Biocybernetics and Biomedical Engineering, vol. 42, no. 1, pp. 27–41, 2022.

[46] A. A. Hammam, H. H. Elmousalami and A. E. Hassanien, "Stacking Deep Learning for Early COVID-19 Vision Diagnosis," in Chapter: Big Data Analytics and Artificial Intelligence Against COVID-19: Innovation Vision and Approach, vol. 78, pp. 297–307, DOI: 10.1007/978-3-030-55258-9_18, 2020.

[47] J. A. ALzubi, B. Bharathikannan, S. Tanwar, R. Manikandan, A. Khanna and C. Thaventhiran, "Boosted Neural Network Ensemble Classification for Lung Cancer Disease Diagnosis," Applied Soft Computing, vol. 80, pp. 579–591, 2019.

[48] G. Li, R. Togo, T. Ogawa and M. Haseyama, "Boosting Automatic COVID-19 Detection Performance with Self-supervised Learning and Batch Knowledge Ensembling," Computers in Biology and Medicine, vol. 158, Article no. 106877, 2023.

[49] N. Habib, M. M. Hasan, M. M. Reza and M. M. Rahman, "Ensemble of CheXNet and VGG-19 Feature Extractor with Random Forest Classifier for Pediatric Pneumonia Detection," SN Computer Science, vol. 1, no. 6, Article no. 359, 2020.

[50] A. B. Godbin and S. G. Jasmine, "A Machine Learning Based Approach for Diagnosing Pneumonia with Boosting Techniques," in Chapter: Machine Intelligence for Smart Applications, Part of the Book Series: Studies in Computational Intelligence, vol. 1105, pp. 145–160, Springer, 2023.

ملخص البحث:

إنّ أمراض الرّئة، مثل فيروس كورونا وذات الرّئة، يمكن أنْ تسبّب تداعيات شديدة، بما في ذلك صعوبات في التّنفُّس، وضعف في عمل الرّئة، وفشل في الجهاز التّنفُّسي، ممّا قد يهدّد الحياة إذا لم يُعالج على الفْور. وقد أثبتت الصُّور الشُّعاعية للصّدْر سرعتها وفاعليتها وجدواها في تشْخيص هذه الأمراض والكَشْف عنها. بالإضافة إلى ذلك، أعطى الذّكاء الاصطناعي من خلال تعلُّم الآلة والتّعلُّم العميق نتائج واعدةً في الكشْف عن العديد من أمراض الرّئة، ومنها فيروس كورونا وذات الرّئة. وقد أسهمت قدرةُ هذه التكنولوجيا على التّعامل مع مجموعات البيانات الضّخمة وتحليلها في الحدّ من انتشار هذه الأمراض وكان لها دور بارز في تقدُّم البحث العلمي في العديد من التّخصُّصات الطّبّية والبيولوجية.

في هذا البحث، نستخدم مجموعةً من تقنيات توحيد الأداء عن طريق خوارزميات مختلفة لتحسين أداء نماذج التّصنيف في الكشْف عن الإصابة بفيروس كورونا وذات الرّئة. وبالتّحديد، فقد ركّزنا على إيجاد نماذج مجمَّعة تقوم على توحيد الأداء وتستند على الشّبكات العصبية الالتفافية ومحوّلات الرّؤية. وكان هدفنا تحديد أفضل تلك النّماذج القائمة على توحيد الأداء في تشْخيص أمراض الرّئة، وذلك من خلال تقييم قدرة النّماذج المستخدمة على التّصنيف الصّحيح للصُّور الشّعاعية للصّدر إلى صُوَرٍ تدلُّ على الإصابة بفيروس كورونا، وأخرى تدلُّ على الإصابة بذات الرّئة، وثالثة طبيعية تشير ألى الخُلوّ من أمراض الرّئة.

ويمكن القول إنّ تقنيات توحيد الأداء تعْمل على تحسين دقّة التّشخيص للنّماذج المستخدمة في الكشْف عن أمراض الرّئة، القائمة على تعلُّم الآلة.

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 10, No. 04, December 2024.

DENOISING DIFFUSION PROBABILISTIC MODEL WITH WAVELET PACKET TRANSFORM FOR FINGERPRINT GENERATION

Li Chen and Huah Yong Chan

(Received: 29-Jun.-2024, Revised: 9-Aug.-2024, Accepted: 12-Aug.-2024)

ABSTRACT

The majority of contemporary fingerprint synthesis is based on the Generative Adversarial Network (GAN). Recently, the Denoising Diffusion Probabilistic Model (DDPM) has been demonstrated to be more effective than GAN in numerous scenarios, particularly in terms of diversity and fidelity. This research develops a model based on the enhanced DDPM for fingerprint generation. Specifically, the image is decomposed into sub-images of varying frequency sub-bands through the use of a wavelet packet transform (WPT). This method enables DDPM to operate at a more local and detailed level, thereby accurately obtaining the characteristics of the data. Furthermore, a polynomial noise schedule has been designed to replace the linear noise strategy, which can result in a smoother noise-addition process. Experiments based on multiple metrics on the datasets SOCOFing and NIST4 demonstrate that the proposed model is superior to existing models.

KEYWORDS

Diffusion model, Fingerprint, Image processing, Wavelet packet transform.

1. INTRODUCTION

The singularity and immutability of fingerprints render them an indispensable biometric trait of humans, which is extensively utilized in a multitude of contexts including identity authentication, criminal investigation, medical research and so forth [1]–[3]. Nevertheless, the establishment of such a fingerprint system necessitates a substantial quantity of fingerprint samples and direct fingerprint collection is constrained by equipment, environmental or legal considerations. In response to this challenge, the advent of computer vision has led to the emergence of synthetic fingerprint technology. However, the intricacies of fingerprints are considerable, encompassing a multitude of diverse pattern types, including whorls, loops and arches [4].

Most early model-based methods typically synthesize or reconstruct fingerprints based on minutiae templates or shallow model. While these methods have been demonstrated to be effective in synthesizing fingerprint structure and texture [5]–[8], they exhibit low automation levels and poor model adaptability and scalability. Furthermore, deep learning-based generative models have gradually become the dominant models for fingerprint generation, employing more complex network structures, such as auto-encoders and GANs [9]–[13]. However, images generated by VAEs are often quite blurry and the shortcomings of GAN models, such as mode collapse and difficulty in training, cannot be ignored either.

The recent emergence of Denoising Diffusion Probabilistic Models (DDPMs) [14] has ushered in a novel technical approach to image synthesis. This model is based on the Markov chain, which gradually introduces noise to the data in the forward stage until the data is corrupted and becomes completely Gaussian noise. In the reverse stage, the Gaussian noise is then restored to the original data. The model must ensure that the reverse Markov chain closely resembles the forward process during optimization. It has demonstrated remarkable efficacy in the generation of images, videos and other forms of data, particularly in the domain of image synthesis, as evidenced by its performance in DALL-E 2 [15] and Stable Diffusion [16].

This study proposes a novel fingerprint-generation method based on an enhanced model of DDPM. Specifically, the original image is decomposed into sub-images of varying frequency bands through wavelet packet transform. These sub-images contain information of different scales and directions, which enables DDPM to operate at a more local and detailed level, thereby improving the stability of

L. Chen and H.Y.Chan are with School of Computer Sciences, Universiti Sains Malaysia, 11800, Pulau Pinang, Malaysia. Emails: chenli0926@student.usm.my and hychan@usm.my

training and the quality of generated images. Moreover, in consideration of the potential for the linear noise schedule in the original DDPM to result in the corruption of data too quickly in the forward phase, this study devised a polynomial noise schedule to facilitate a more gradual and less disruptive noise-addition process, which prevents an abrupt change in noise level. Experimental results have demonstrated the effectiveness of this approach in both SOCOFig [17] and NIST Special Database 4 [18].

2. RELATED WORK

Inspired by non-equilibrium statistical physics, [19] designed a parameterized Markov-chain model that slowly destroys data in the forward phase and gradually recovers data in the backward phase which is highly flexible and tractable. DDPM model further extended this idea to high quality tasks. The forward process commences with the real distribution of data $q(x_0)$ and incorporates a minimal quantity of Gaussian noise at each stage, gradually transforming the data into pure noise. The whole process can be represented by $q(x_t|(x_{t-1})$. In the generation process, DDPM starts with a standard Gaussian noise sample x_t, gradually removes the noise through the reverse process and finally generates a sample x_0, which can be defined as $p_\theta((x_{t-1}|(x_t)$.

2.1 The Principle of DDPM

In the forward process, let $x_0, x_1 \ldots\ldots x_n$ represent the gradual addition of Gaussian noise to an image until the image becomes completely noisy. This process can be described as shown in Figure 1.

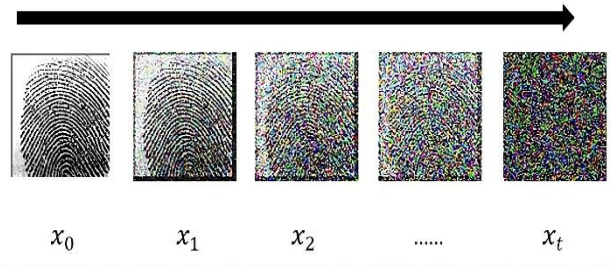


Figure 1. The forward process of DDPM.

It should be noted that the degree of noise added in each step is not uniform. It is controlled by weight β and its value must be gradually increased. For example, it normally ranges from 0.0001 to 0.02. The actual noise added is represented by α, which is defined as $\alpha_t = 1 - \beta_t$ to represent the weight of the noise at step t. Since the image obtained at each step depends on the image from the previous step plus the noise $z \sim N(0, 1)$ of the current step, that is the model relies on a Markov chain, the image at any step can be obtained by $x_t = \sqrt{\alpha_t} x_{t-1} + \sqrt{1 - \alpha_t} z_t$. As the diffusion process continues, the noise content of the image will increase until it reaches a point where the entire image is comprised of noise. Due to the recursive nature of the model, the diffusion process does not occur in discrete steps during training. Instead, the model directly calculates the relationship between the original image and the noise image at any given step. This relationship is represented by the cumulative product of all steps, denoted by $\overline{\alpha_t}$. In the reverse process, the model must restore the noise to the image (Figure 2), Although it is challenging to determine $p_\theta(x_{t-1}|x_t)$ directly, x_0 and x_t are known in the forward process, which is available to calculate posterior probability $q(x_{t-1}|x_t, x_0)$ to estimate the mean and variance of reverse-process distribution. After Bayesian transformation, the goal can be transformed into calculating the distribution (2), (3), (4).

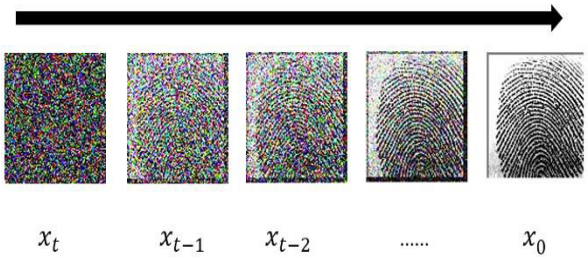


Figure 2. The reverse process of DDPM.

$$q(x_{t-1}|x_t, x_0) = q(x_t|x_{t-1}, x_0)\frac{q(x_{t-1}|x_0)}{q(x_t|x_0)} \quad (1)$$

$$q(x_{t-1}|x_0) = \sqrt{\overline{\alpha_{t-1}}}x_0 + \sqrt{1-\overline{\alpha_{t-1}}}z_{t-1} \sim N\left(\sqrt{\overline{\alpha_{t-1}}}x_0, 1-\overline{\alpha_{t-1}}\right) \quad (2)$$

$$q(x_t|x_0) = \sqrt{\overline{\alpha_t}}x_0 + \sqrt{1-\overline{\alpha_t}}z_t \sim N\left(\sqrt{\overline{\alpha_t}}x_0, 1-\overline{\alpha_t}\right) \quad (3)$$

$$q(x_t|x_{t-1}, x_0) = \sqrt{\alpha_t}x_{t-1} + \sqrt{1-\alpha_t}z_t \sim N\left(\sqrt{\alpha_t}x_{t-1}, 1-\alpha_t\right) \quad (4)$$

Since all three distributions are normal distributions, but with different means and variances, according to the normal distribution probability density function, the mean (μ) and variance (σ) of the reverse-process distribution can be obtained:

$$\sigma^2 = \frac{\alpha_t}{\beta_t} + \frac{1}{1-\overline{\alpha_{t-1}}}{x_{t-1}}^2 \quad (5)$$

$$\mu = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{1-\overline{\alpha_t}}\epsilon_\theta\right) \quad (6)$$

Where ϵ_θ is the predicted value of z. Finally, the loss function can be defined as follows:

$$L_{simple}(\theta) = E_{t,x_0,z} \parallel \mathrm{z} - \epsilon_\theta\left(\sqrt{\overline{\alpha_t}}x_0 + \sqrt{1-\overline{\alpha_t}}z, t\right) \parallel^2 \quad (7)$$

2.2 The Improvements of DDPM

Although DDPM has made considerable achievements, it still has shortcomings. [20] found that adding linear noise may cause the $\overline{\alpha_t}$ to approach 0 too quickly, which may damage data information quickly during training. It proposes adding a certain amount of cosine noise in the forward process, which can obtain better log-likelihoods. [21] established a noise signal in a learnable format and incorporated Fourier features into the input of the network, enabling the prediction of noise. These enhancements exceed the performance of autoregressive models. [22] put forth a methodology for dynamically adjusting the noise parameters with the objective of enhancing the denoising capabilities of the model, thereby improving the quality of the synthesized results. [23] introduced an Adversarial Purification approach, which incorporates noise and adversarial images into the forward process. In the reverse process, it is necessary to remove both the noise and the adversarial perturbation, while retaining the main content of the input image. This has the effect of improving the classification task. [24] presented a non-Markovian mechanism that renders the sampling process deterministic in the reverse process. This improvement has led to an improvement in the efficiency of the sampling process.

Some researchers have employed a combination of signal-processing techniques with DDPM. The wavelet transform is a widely utilized signal-processing technique that enables the decomposition of signals into components with varying frequencies and time resolutions. In the field of image processing, the wavelet transform is a valuable tool that can assist in the analysis and processing of information at different scales and frequencies within images. Its applications include tasks, such as image denoising, compression and feature extraction. In a recent study, [25] developed a model that combined wavelet and DDPM for 3D medical scans. This model performed a wavelet transform on the input to predict both wavelet coefficients and noise. [26] posited that the use of wavelet transform can mitigate the long-term inference issue by transferring the image reconstruction task from the spatial domain to the wavelet domain. [27] also used a similar concept, proposing a Stage-by-stage Wavelet Optimization Refinement Diffusion model, which takes wavelet transformation to improve the robustness of the model. The authors further argued that wavelet transform can facilitate the disentanglement of image content and features at varying scales, thereby enhancing the stability of the model-training process. [28] extracted the high-frequency and low-frequency information of the image after wavelet transformation in order to accelerate the training of the model without compromising the quality of the output.

3. DDPM WITH WAVELET PACKET TRANSFORM

This study implemented the following improvements based on DDPM. Firstly, the Wavelet Packet Transform (WPT) was initially introduced in the training stage. In this stage, the original data was transformed into wavelet packets to obtain sub-bands at different scales and frequencies. The diffusion

process was applied to the sub-bands of each scale and frequency to gradually add noise, which helps the model better extract data features and improve training stability. In the sampling stage, the original image was reconstructed using the inverse wavelet packet transform (IWPT). Moreover, a polynomial noise schedule was devised with the objective of reducing abrupt change of noise level. This strategy allows for the addition of noise in a more gradual and smoother manner, thereby reducing the potential for information loss.

3.1 Principle of Wavelet Packet Transform

The wavelet transform is a mathematical tool that employs various wavelets to decompose a signal into low-frequency and high-frequency components including the Morlet, Daubechies and Haar wavelets. Wavelet transform can provide both time and frequency information. It has a wide range of applications in image denoising, image compression, feature extraction and so forth. Wavelet packet transform is a generalization of wavelet transform, allowing further decomposition of the approximate and detailed parts of the signal. It provides finer frequency resolution by recursively decomposing all frequency bands of the signal. Specifically, for the discrete signal $f(t)$, its decomposition by wavelet packet transform can be expressed as follows:

$$W_{j,k}(t) = \sum_{t} f(t)\psi_{j,k}(t) \tag{8}$$

In this context, $\psi_{j,k}(t)$ represents the function of the wavelet packet transform, with j and k denoting the scale and position parameters, respectively. The reconstruction of the inverse wavelet packet transform can be defined as follows:

$$f(t) = \sum_{j} \sum_{k} W_{j,k}\psi_{j,k}(t) \tag{9}$$

The characteristics of WPT are employed to perform multi-scale processing on the model. Multi-scale analysis facilitates the capture of structural information within the data, thereby enhancing the quality of the denoising and reconstruction process. The addition of noise to different frequency components independently serves to render the distribution of noise more reasonable, which in turn improves the quality of the generated data.

3.2 Wavelet Packet Transform on DDPM

In the forward process, let $x \in \mathbb{R}^{D\times H\times W}$ denote the input data. During the experiment, the three-level wavelet packet transform of Haar was employed on the input data, with l and h indicating low-frequency and high-frequency sub-bands, respectively. Following the application of the wavelet packet transform to the first layer, four sub-images can be derived, which are: x_{ll}^1, x_{lh}^1, x_{hl}^1, x_{hh}^1, all of which are elements of $\mathbb{R}^{\frac{D}{2}\times\frac{H}{2}\times\frac{W}{2}}$. Subsequently, the two dimensions of x_{ll}^1 and x_{lh}^1 are concatenated in the width direction and the two dimensions of x_{hl}^1 and x_{hh}^1 are concatenated in the height direction. Finally, the two aforementioned structures are concatenated to obtain the matrix $x^1 \in \mathbb{R}^{D\times H\times W}$. As with a binary tree, it can be demonstrated that 16 sub-images can be obtained in the second layer and 64 sub-images in the third layer. Similar operations are then performed on the second and third layers to obtain x^2 and x^3, which are subsequently spliced from each level to finally yield $y \in \mathbb{R}^{3(D\times H\times W)}$, which is then used for DDPM processing. The complete training and sampling process of the model is as follows:

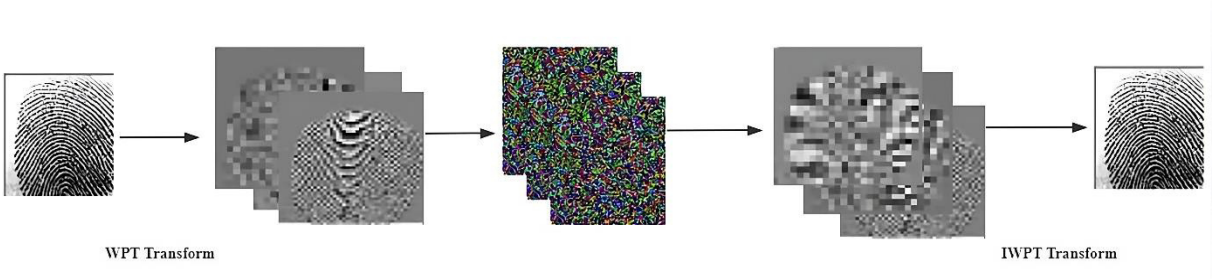


Figure 3. The pipeline of WPT diffusion.

Algorithm 1 Training

Repeat

$y_0 = WPT(x_0)$

T~Uniform (1,…,T)

$\epsilon \sim N(0,1)$

$q(y_t|y_{t-1}) = N\left(y_t, \sqrt{\alpha_t} y_{t-1}, (1-\alpha_t)I\right)$

Take gradient descent step on $\nabla_\theta \parallel \epsilon - \epsilon_\theta(y_t, t) \parallel^2$

Until converge

Algorithm 2 Sampling

$y_t \sim N(0, I)$

for t = T,…,1 do

z $\epsilon \sim N(0, I)$

$p_\theta = \left(y_{t-1} | y_t, z, \epsilon_\theta(y_t, t)\right)$

$x_0 = IWPT\ (y_0)$

return x_0

During the training process, the input data is transformed by WPT into several sub-bands and all sub-band signals are spliced and trained in the WPT domain rather than in the original pixel domain. In the sampling process, the distribution of y_t is first determined to be $N(0, I)$ and the entire reverse process is then applied to obtain all sub-images. The image y_0 is finally reconstructed by performing an inverse wavelet packet transform (IWPT).

3.3 Polynomial Noise Schedule

Linear noise schedules may bring abrupt changes in noise level during training, which causes information disruption too quickly during training, especially for images with lower resolution (less than or equal to 64× 64) [20]. To avoid this phenomenon, this research designs a polynomial noise schedule, where this noise strategy is smoother in the process of adding, especially at $t = 0$ and $t = T$. It is milder in noise changes than the linear schedule and cosine schedule, which helps the stability of training and avoids instability of training caused by sudden noise changes.

$$\beta_t = \beta_0 + (\beta_T - \beta_0)(\frac{t}{T})^p \tag{10}$$

As Equation (10) shows, let β be the parameter to control the noise level, where β ranges form β_0 to β_T, t and T denote the current step and max step. By adjusting the power parameter p, the growth rate of β can be controlled, thereby affecting the smoothness of the cumulative product of $\overline{\alpha_t}$. p is the hyperparameter to adjust the smoothness of function. $p = 2$ is default setting during training, which is experimentally workable. Figure 4 shows that this noise schedule is smoother than the linear strategy and cosine strategy at the beginning and end of training.

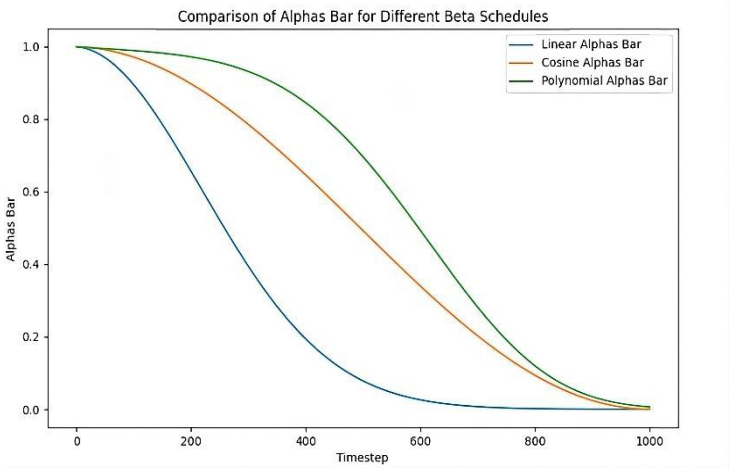


Figure 4. Compassion between different noise schedules.

3.4 Theoretical Analysis

In comparison to the conventional DDPM model, the methodology presented in this study offers enhanced capabilities in several key areas. Initially, the wavelet packet transform is capable of decomposing an image into sub-images of varying frequencies and directions, thereby encompassing the multi-scale characteristics of the original image. Moreover, it enables the separation of high-frequency and low-frequency components within an image, thereby facilitating the processing of these components in isolation and the reduction of blurring effects in the generated image. During the diffusion process, the model is capable of processing sub-images of varying frequencies in a manner that preserves and restores details and coarse information in a more optimal manner. In addition, the wavelet packet transform enables the separation of different image features, thereby enhancing the model's resilience to noise and other disturbances. Sub-images of varying frequency bands can be processed independently during the diffusion process, reducing mutual interference between different frequency bands. This improves the stability and quality of the generated image. Furthermore, polynomial noise schedule ensures a smoother way to add noise to each sub-image, which will contribute to the stability of model training.

4. EXPERIMENTS

The proposed model is evaluated with 200-epoch training on two distinct datasets: SOCOFing and NIST4 which are mentioned before. During the experiment, the architecture is based on U-Net included two attention blocks with a learning rate of 0.0001, β ranging from 0.0004 to 0.02 and a maximum step T of 1000. Finally, images with a resolution of 64×64 are generated.

4.1 Datasets

The Sokoto Coventry Fingerprint Dataset (SOCOFing) is a biometric fingerprint database created for academic-research purposes. This dataset comprises 6,000 fingerprint images with a resolution of 96×103 pixels, captured from 600 subjects of African descent. Each image is accompanied by a set of attributes, including gender, hand and finger name labels, as well as a sub-set of altered fingerprint images. This experiment employs all 6,000 real fingerprint images as training data.

NIST Special Database 4 (NIST 4) is a tool for evaluating fingerprint systems and contains 2000 8-bit grayscale fingerprint pairs, totaling 4000 images with 512×512 resolution. The database may be employed for the purposes of algorithm development, system training and testing. The experiment utilizes all 4,000 images as training data.

4.2 Progressive Generation

First, to verify the effectiveness of the proposed model, a sample was taken every 100 steps at T=1000 and 10 images were generated to observe the reverse diffusion process. As shown in Figure 5, which illustrates the reverse process of the proposed model, the image starts with noise, then evolves to the fingerprint outline and finally to a recognizable fingerprint texture. This result is indicative of the effectiveness of the model proposed in this research.

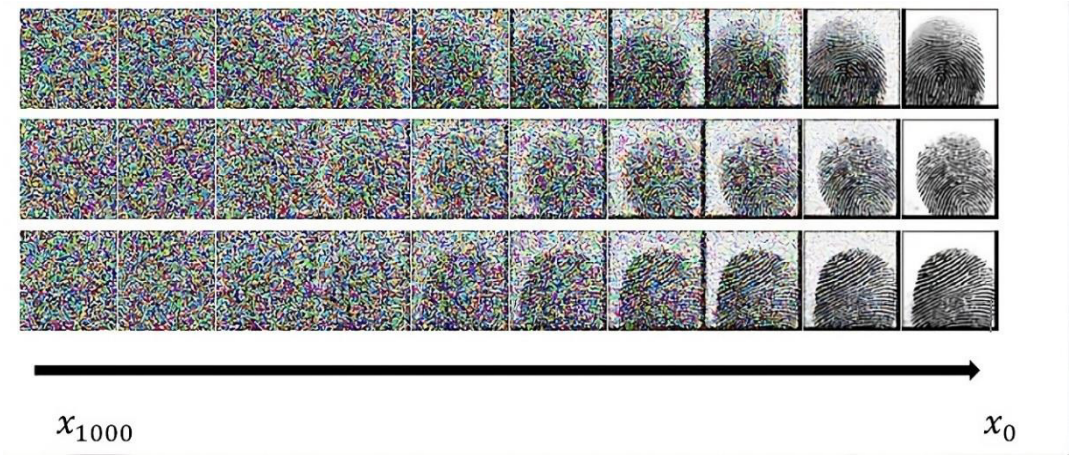


Figure 5. The reverse process of the proposed model.

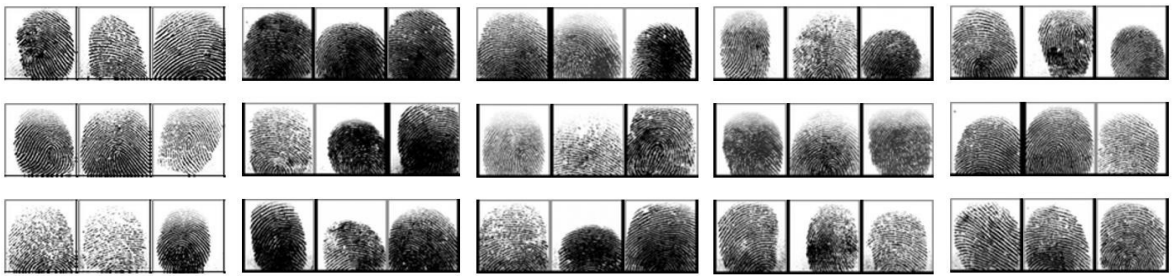
4.3 Results of Quality Evaluation

In the quality assessment, this sub-section firstly employs the Frechet Inception Distance (FID) [29] and Inception Score (IS) [30] as evaluation metrics on both SOCOFing and NIST4. FID and IS are two of the most commonly used metrics for evaluating the quality of images generated by generative models. Both require the utilization of pre-trained Inception v3 networks. The following models are employed for comparative purposes: DDPM (Vanilla) [14], DDPM (Wavelet) [25] and FingerGAN [12]. The results are presented in Table 1.

Table 1. The comparison between different models on IS and FID.

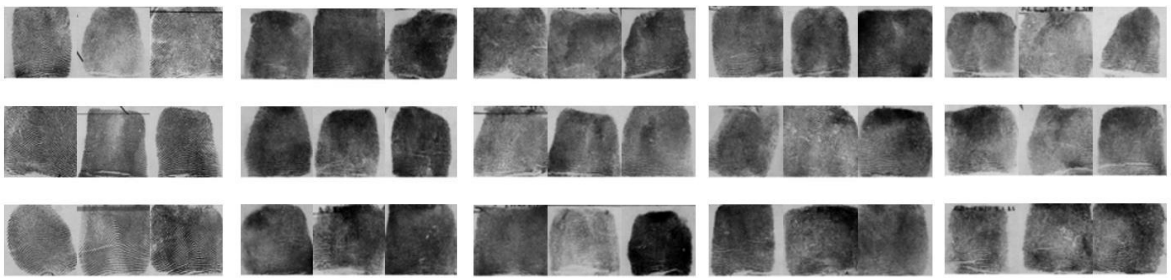
Model	SOCOFing		NIST4	
	FID	**IS**	**FID**	**IS**
DDPM(Vanilla) [14]	64.61	2.26 $\pm$0.16	62.71	2.69 $\pm$0.19
DDPM (Wavelet) [25]	62.96	2.33 $\pm$0.11	60.53	**2.73 $\pm$0.21**
FingerGAN [12]	70.35	2.12 $\pm$0.14	69.21	2.33 $\pm$0.12
Proposed Model	**61.64**	**2.42 $\pm$0.13**	**59.65**	2.72 $\pm$0.11

Table 1 demonstrates that the proposed model outperforms the other three models in terms of both FID and IS in the majority of scenarios. Nevertheless, the IS of DDPM (Wavlet) is marginally superior to the proposed model on the NIST4 dataset. This indicates that the proposed model is capable of generating high-quality and diverse results. Furthermore, this sub-section presents a comparative analysis of the images generated by all models, as illustrated in Figures 6 and 7. The data represented by the labels 1, 2, 3, 4, 5 in Figures 6 and 7 is real data, data generated by FingerGAN, data generated by DDPM, data generated by DDPM (wavelet) and data generated by the proposed model, respectively. In terms of visual quality, the four models demonstrate satisfactory performance on SOCOFing. However, the results produced by Finger GAN are not sufficiently diverse and the model proposed in this paper is more effective at generating fingerprint texture and structure. Furthermore, given the high resolution of the NIST4 original data, reaching at 512×512 resolution, it is not evident that there is a discernible visual difference between the four models at the resolution of 64×64.



(1) (2) (3) (4) (5)

Figure 6. The generated results from each model on SOCOFing.



(1) (2) (3) (4) (5)

Figure 7. The generated results from each model on NIST4.

In addition, as both FID and IS depend on the pretrained model, this sub-section also employs SSIM and MS-SSIM to assess the performance of the model. SSIM (Structural Similarity Index) is an index that gauges the similarity of two images. It considers the image's brightness, contrast and structure. MS-SSIM (Multi-scale SSIM) is an extension of SSIM that assesses the structural similarity of images at different scales. For the SOCOFing dataset, the generated original images were resized to 64×64 24-bit width. For the NIST4 data set, the generated original images were resized to 64×64 8-bit width. Finally,

the Average SSIM and Average MS-SSIM were calculated, as shown in Table 2.

As can be seen from Table 2, in general, the value of MS-SSIM is lower than that of SSIM. This discrepancy may be attributed to the fact that MS-SSIM considers the structural similarity of images at multiple scales, thereby providing a more comprehensive image-quality assessment. It can be observed that the proposed model exhibits better performance in both SSIM and MS-SSIM, particularly in SSIM on NIST4, where its score is considerably higher than those of other models. This may indicate that the proposed model is more adept at maintaining image quality. In addition, The MS-SSIM metric also demonstrates that the proposed model exhibits certain advantages when dealing with scenes involving image scaling or multi-scale characteristics.

Table 2. The comparison between different models on SSIM and MS-SSIM.

Model	SOCOFing		NIST4	
	SSIM	MS- SSIM	SSIM	MS- SSIM
DDPM(Vanilla) [14]	0.239	0.215	0.441	0.281
DDPM (Wavelet) [25]	**0.261**	0.227	0.452	0.262
FingerGAN [12]	0.227	0.203	0.346	0.272
Proposed Model	0.253	**0.239**	**0.495**	**0.295**

4.4 Ablation Study

In order to ascertain the effect of each modification to the model on its overall performance, an ablation study is conducted in this sub-section. The present study proposes the utilization of the Wavelet Packet Transformation (WPT) technique to decompose the original data into a number of sub-bands. Prior research has demonstrated that the application of wavelet transformation has resulted in a more stable training process. Wavelet packet transformation (WPT) is an extension of wavelet transformation (WT) offering a more detailed frequency decomposition. In comparison to WT, WPT further decomposes all components (including low-frequency and high-frequency) at each decomposition level, thus facilitating the extraction of more accurate features, which in turn improves the stability and quality of image generation. By training at 200 epochs, the loss changes between vanilla DDPM and the DDPM with WPT are shown in Figure 8. Obviously, these two models converge at approximately 20 epochs, but the loss of vanilla DDPM fluctuates. In contrast, the loss of DDPM with WPT is more stable.

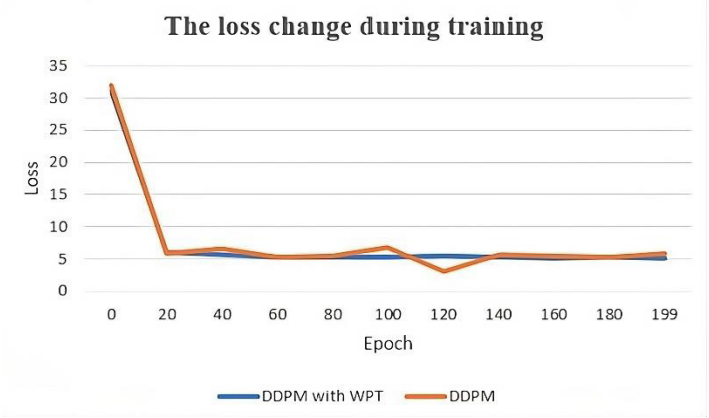


Figure 8. Loss comparison between DDPM and DDPM with WPT.

According to Table 3, the proposed improvements have led to an enhancement in the model performance, with WPT exhibiting the most notable improvement. Although the contribution of Polynomial noise schedule is not as great as that of WPT, it is demonstrably superior to the original linear schedule.

Table 3. The ablation study of each component.

Model	SOCOFing		NIST4	
	FID	IS	FID	IS
DDPM (Linear)	64.61	2.26 $\pm$0.16	62.71	2.69 $\pm$0.19
DDPM (Linear +WPT)	61.77	2.37 $\pm$0.21	60.22	2.71 $\pm$0.13
DDPM (Polynomial)	63.89	2.29 $\pm$0.14	61.98	2.70 $\pm$0.14
DDPM (Polynomial +WPT)	**61.64**	**2.42 $\pm$0.13**	**59.65**	**2.72 $\pm$0.11**

5. CONCLUSION AND FUTURE RESEARCH

This study proposes an improved version of DDPM for the task of fingerprint generation. Firstly, wavelet packet transform is used during training, which allows the model to better extract data features on different sub-bands. Secondly, a better noise schedule is developed to make noise addition smoother during the training process. The results of the experiments demonstrate that the proposed model exhibits superior performance compared to previous models. Nevertheless, the inference time remains comparable to that of the original DDPM, which represents a potential avenue for future research. Besides, the current equipment limits the resolution of the images generated by this model to 64×64, so the objective is to apply this model to a higher resolution in the future. It would be beneficial to explore the potential of this model for generating text and video as well.

REFERENCES

[1] J. Priesnitz et al., "An Overview of Touchless 2D Fingerprint Recognition," EURASIP Journal on Image and Video Processing, vol. 2021, no. 1, p. 8, Dec. 2021.

[2] J. Kolberg et al., "COLFISPOOF: A New Database for Contactless Fingerprint Presentation Attack Detection Research," Proc. of the 2023 IEEE/CVF Winter Conf. on Applications of Computer Vision Workshops (WACVW), pp. 653–661, Waikoloa, HI, USA, 2023.

[3] E. Prabakaran and K. Pillay, "Nanomaterials for Latent Fingerprint Detection: A Review," Journal of Materials Research and Technology, vol. 12, pp. 1856–1885, May 2021.

[4] M. Kücken M and A. Newell, "Fingerprint Formation," Journal of Theoretical Biology, vol. 235, no. 1, pp. 71–83, Jul. 2005.

[5] R. Cappelli et al., "Synthetic Fingerprint-image Generation," Proc. of the 15th Int. Conf. on Pattern Recognition (ICPR-2000), vol. 3, pp. 471–474, Barcelona, Spain, 2000.

[6] J. L. Araque et al., "Synthesis of Fingerprint Images," Proc. of the 2002 Int. Conf. on Pattern Recognition, vol. 2, pp. 422–425, DOI: 10.1109/ICPR.2002.1048329, Quebec City, QC, Canada, 2002.

[7] J. Feng and A. K. Jain, "Fingerprint Reconstruction: from Minutiae to Phase," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 2, pp. 209–223, 2011.

[8] T. Uz et al., "Minutiae-based Template Synthesis and Matching for Fingerprint Authentication," Computer Vision and Image Understanding, vol. 113, no. 9, pp. 979–992, Sep. 2009.

[9] M. Attia, "Fingerprint Synthesis *via* Latent Space Representation," Proc. of the 2019 IEEE Int. Conf. on Systems, Man and Cybernetics (SMC), pp. 1855–1861, Bari, Italy, 2019.

[10] O. Striuk and Y. Kondratenko, "Adaptive Deep Convolutional GAN for Fingerprint Sample Synthesis," Proc. of the 2021 IEEE 4th Int. Conf. on Advanced Information and Communication Technologies (AICT), pp. 193–196, DOI: 10.1109/AICT52120.2021.9628978, Lviv, Ukraine, 2021.

[11] P. Schuch et al., "De-convolutional Auto-encoder for Enhancement of Fingerprint Samples," Proc. of the 2016 6th Int. Conf. on Image Processing Theory, Tools and Applications (IPTA), pp. 1–7, Oulu, Finland, DOI: 10.1109/IPTA.2016.7821036, 2016.

[12] S. Minaee and A. Abdolrashidi, "Finger-GAN: Generating Realistic Fingerprint Images Using Connectivity Imposed GAN," arXiv, [Online], Available http://arxiv.org/abs/1812.10482, Dec. 2018.

[13] S. Seidlitz et al., "Generation of Privacy-friendly Datasets of Latent Fingerprint Images Using Generative Adversarial Networks," Proc. of the 16th Int. Joint Conf. on Computer Vision, Imaging and Computer Graphics Theory and Applications, pp. 345–352, DOI: 10.5220/0010251603450352, 2021.

[14] J. Ho et al., "Denoising Diffusion Probabilistic Models," Proc. of the 34th Int. Conf. on Neural Information Processing Systems, pp.6840-6851, Vancouver, BC, Canada, 2020.

[15] A. Ramesh et al., "Hierarchical Text-conditional Image Generation with CLIP Latents," arXiv, [Online], Available http://arxiv.org/abs/2204.06125, Apr. 2022.

[16] R. Rombach et al., "High-resolution Image Synthesis with Latent Diffusion Models," Proc. of the 2022 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 10674–10685, New Orleans, LA, USA, 2022.

[17] Y. I. Shehu et al., "Sokoto Coventry Fingerprint Dataset," arXiv, [Online], Available https://arxiv.org/abs/1807.10609, 2018.

[18] C. Watson, NIST Special Database 4, NIST 8-Bit Gray Scale Images of Fingerprint Image Groups, World Wide Web-Internet and Web Information Systems, [Online], Available: https://www.nist.gov/publications/nist-special-database-4-nist-8-bit-gray-scale-images-fingerprint-image-groups, 2008.

[19] J. Sohl-Dickstein et al., "Deep Unsupervised Learning Using Nonequilibrium Thermodynamics," Proc. of the 32nd Int. Conf. on Machine Learning, vol. 37, pp.2256 - 2265, 2015.

[20] A. Nichol and P. Dhariwal, "Improved Denoising Diffusion Probabilistic Models," Proc. of the 38th Int. Conf. on Machine Learning, vol 139, pp.8162-8171, 2021.

[21] D. P. Kingma et al., "Variational Diffusion Models," Proc. of the 35th Int. Conf. on Neural Information

Processing Systems, pp. 21696-21707, DOI: 10.5555/3540261.3541921, 2021.
[22] R. San-Roman et al., "Noise Estimation for Generative Diffusion Models," arXiv, [Online], Available http://arxiv.org/abs/2104.02600, Sep. 2021.
[23] J. Wang et al., "Guided Diffusion Model for Adversarial Purification," arXiv, [Online], Available http://arxiv.org/abs/2205.14969, Jun. 2022.
[24] J. Song et al., "Denoising Diffusion Implicit Models," Proc. of the 2021 Int. Conf. on Learning Representations, Vienna, Austria, [Online], Available https://openreview.net/forum?id=St1giarCHLP, 2021.
[25] P. Friedrich et al., "WDM: 3D Wavelet Diffusion Models for High-Resolution Medical Image Synthesis," arXiv, [Online], Available http://arxiv.org/abs/2402.19043, Feb. 2024.
[26] Y. Huang et al., "WaveDM: Wavelet-based Diffusion Models for Image Restoration," IEEE Transactions on Multimedia, vol. 26, pp. 7058-7073, DOI: 10.1109/TMM.2024.3359769, 2024.
[27] K. Xu et al., "Stage-by-stage Wavelet Optimization Refinement Diffusion Model for Sparse-View CT Reconstruction," IEEE Trans. Med. Imaging, DOI: 10.1109/TMI.2024.3355455, 2024.
[28] H. Phung et al., "Wavelet Diffusion Models Are Fast and Scalable Image Generators," Proc. of the 2023 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 10199–10208, Vancouver, BC, Canada, 2023.
[29] N. B. Bynagari, "GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium," Asian Journal of Applied Sciences & Eng., vol. 8, no. 1, pp. 25–34, Apr. 2019.
[30] S. Barratt and R. Sharma, "A Note on the Inception Score," Proc. ICML 2018 Workshop on Theoretical Foundations and Applications of Deep Generative Models, Stockholm, Sweden, [Online], Available https://drive.google.com/file/d/1y--gMfGtjlNcQXTPm-vLhuqlUbh19nF9/view, 2018.

ملخص البحث:

تستند غالبية تقنيات تركيب بَصَمات الأصابع على استخدام الشّبكات التّوليدية التّنافسية (GANs). وحديثاً، أثبت النّموذج الاحتمالي للانتشار القائم على إزالةِ الضّجيج (DDPM) أنّه أكثرُ فاعليةً من ذلك النّوع من الشّبكات في العديد من السّيناريوهات، لا سيما فيما يتعلّق بالتنوُّع والوُضوح.

في هذا البحث، نعملُ على تطوير نموذجٍ محسَّنٍ يقومُ على تقنية (DDPM) لتوليد بَصَمات الأصابع. وبالتحديد، فإنّه يتمّ تفكيك الصّورة إلى صُوَر فرعية بنطاقاتٍ تردُّديةٍ فرعيةٍ مختلفة، وذلك باستخدام طريقة تحويل حُزَم المويْجات (WPT) التي تجعل تقنية (DDPM) تعمل عند مستوى أكثر محلّيةً وتفصيلاً، مُسترجعةً بذلك خصائص البيانات على نحوٍ دقيقٍ. من ناحية أخرى، تمّ استخدام تقنيةٍ جديدةٍ للتّعامل مع الضّجيج لتحلَّ محلّ الاستراتيجية الخطّية لمعالجة الضّجيج، بحيث تكون عملية التّعامُل مع الضّجيج أكثر سلاسة.

وقد برهنت التّجاربُ على عددٍ من مَجْموعات البيانات المتعلّقة ببَصَمات الأصابع أن النّموذج المقترح في هذا البحث يتفوَّقُ على عددٍ من النّماذج المشابهة الواردة في أدبيات الموضوع.

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 10, No. 04, December 2024.

NOVEL MULTI-CHANNEL DEEP LEARNING MODEL FOR ARABIC NEWS CLASSIFICATION

Imad Jamaleddyn[1], Rachid El Ayachi[1] and Mohamed Biniz[2]

(Received: 4-Jul.-2024, Revised: 11-Aug.-2024 and 26-Aug.-2024, Accepted: 31-Aug.-2024)

ABSTRACT

In the era of digital journalism, the classification of Arabic news presents a significant challenge due to the complex nature of the language and the vast diversity of content. This study introduces a novel multi-channel deep-learning model, CLGNet, designed to enhance the accuracy of Arabic-news categorization. By integrating Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRUs), the proposed model effectively processes and classifies Arabic-text data. Extensive experiments were conducted on multiple datasets, including CNN, BBC and OSAC, where the model achieved outstanding accuracy and robustness, outperforming existing methods. The findings underscore the effectiveness of our hybrid model in addressing the challenges of Arabic-text classification and its potential applications in automated news categorization systems.

KEYWORDS

Convolutional neural networks (CNNs), Long Short-Term Memory (LSTM), Gated Recurrent Units (GRUs), Word embedding, Arabic-text classification.

1. INTRODUCTION

In recent times, digital data has become indispensable in every aspect of modern life, significantly impacting the quality of organizations and financial environments. The focus on data and information quality has intensified in both business and academic realms [1]. The news industry holds particular significance as a purveyor of quality information, essential for the success of any news organization [2].

However, like many other industries, journalism has been profoundly affected by the digital revolution, with online news consumption surpassing print media in recent years. Blogs, social media and online-only newspapers have become increasingly important sources of news, particularly for younger demographics. According to the latest Digital News Report from the Reuters Institute, while a percentage of 32% of the Flemish population still relies on print media, a staggering 78% now obtains news online. Online news consumption has remained stable over the past five years, while newspaper consumption has experienced a significant decline [3].

Consequently, scientists and researchers have endeavored to automate the news-classification process, prompting the development of strategies to identify similarities among news articles and categorize them automatically. This involves categorizing unknown texts to extract their meaning. Formally, this task involves assigning categories to a set of texts represented by x = x1, x2..., xn, labeled with values from a set of categories represented by l = l1, l2..., ln. A classification model is trained using a training dataset, establishing a relationship between features and class labels. The unidentified category of a text can then be determined using this trained model. However, manually completing this task for unannotated documents is labor-intensive and time-consuming. Text classification has found success in numerous fields, including sentiment analysis, information retrieval, spam-mail detection and more [4].

Deep-learning algorithms, particularly in Natural Language Processing (NLP) and text categorization, have gained prominence for their ability to enhance efficiency in maintenance and refurbishment projects. These algorithms, such as Artificial Neural Networks (ANNs), Deep Neural Networks (DNNs), Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Deep Belief

1. I. Jamaleddyn and R. El Ayachi are with Department of Computer Science, Faculty of Sciences and Technics, Moulay Slimane University, Beni Mellal, Morocco. Emails: i.jamaleddyn@gmail.com and rachid.elayachi@usms.ma
2. M. Biniz is with Department of Mathematics and Computer Science, Polydisciplinary Faculty, Moulay Slimane University, Beni Mellal, Morocco. Email: mohamedbiniz@gmail.com

Networks (DBN) and Long Short-Term Memory (LSTM), employ multi-stage, non-linear processing models to abstractly represent data. Recent advancements in text processing, machine learning, training datasets and processing power have bolstered the field's progress [5].

Given the limited resources for Arabic-text classification and the burgeoning amount of Arabic data on the internet, there's a pressing need for systems to manage this significant volume of data. This work proposes a hybrid approach to automate Arabic-news categorization, integrating LSTM, CNNs and gated recurrent units (GRUs) to train a classifier across various news datasets. Natural-language processing (NLP) techniques are employed for processing Arabic texts. The subsequent sections of this article detail the analysis of text classification using deep-learning models, the comprehensive strategy and methodology, discussion and experimental results on popular datasets and finally, conclusions and future-research perspectives.

2. LITERATURE REVIEW

Machine learning and deep learning have significantly advanced natural-language processing (NLP) across various applications, including machine translation [6]-[7], entity resolution [8]-[10], sentiment analysis [11]-[12], question-answering systems [13]-14] and Arabic-news classification [15]-[16].

In recent years, the field of Arabic-news classification has seen considerable advancements, yet it remains fraught with challenges that impede the efficacy of these systems. The primary issue at the heart of this research is the inherent complexity of the Arabic language, characterized by its rich morphology, diverse dialects and unstructured nature. This linguistic complexity poses significant hurdles in natural-language processing (NLP) [17], particularly in the domain of feature extraction and model training. Additionally, there is a notable research gap concerning the evaluation and enhancement of Arabic-specific models, in the context of fake-news detection—a critical area given the rise of misinformation on social-media platforms. The following literature review delves into the current methodologies and innovations aimed at overcoming these challenges, exploring various approaches to feature augmentation, supervised-learning and deep-learning techniques in Arabic NLP.

To address feature sparsity, Zhang et al. proposed augmenting features in non-negative matrix factorization with term (T) and word (W) sets [18]. They introduced clustering indicators to enhance the text-word clustering process, achieving significant improvements over Word2vec and Character-level CNN in testing on diverse datasets, including Twitter sports.

Ameur et al. explored supervised-learning approaches for Arabic-text categorization, integrating static, dynamic and fine-tuned word embeddings with RNNs and CNNs [19]. Their models, tested on the OSAC dataset, demonstrated remarkable effectiveness and performance enhancements. By amalgamating Convolutional Neural Networks and Bidirectional General Recurrent Units, their methods significantly out- performed standard CNN and RNN models, boosting the F-score for Arabic text categorization by 98.61 percentage points.

Bdeir and Ibrahim delved into Arabic-tweet classification using two primary deep-learning approaches, CNN and RNN approaches [20]. Leveraging the Twitter API, they amassed 160,870 Arabic tweets spanning various topics, such as criminal accidents, entertainment, sports and technology. The dataset was divided into 90% (144K tweets) for training and validation and 10% (16K tweets) for testing. Despite feature sparsity in short texts, they found that all deep-learning models performed comparably, achieving a macro-F1 accuracy of 90.1%. CNN, RNN-LSTM and RNN-GRU approaches yielded results ranging from 92.71% to 92.95%.

Hassanein et al. [21] presented a model specifically designed to enhance feature selection for Arabic text classification. They address the challenge posed by the unstructured nature of Arabic text, which complicates machine processing. The authors proposed a multi-step feature-selection approach utilizing the Al-Khaleej-2004 corpus, which encompasses a broad range of news categories. Their methodology begins with pre-processing to extract highly weighted terms that represent the documents' content. This is followed by several steps aimed at refining the feature set. The effectiveness of their feature selection method is evaluated using four classifiers: Naïve Bayes (NB), Decision Tree, CART and KNN classifiers. The study, conducted using WEKA and MATLAB, compares the classifiers based on precision, recall, F-measure and accuracy. Among the classifiers tested, CART classifier demonstrated the best performance, while KNN classifier was found to be the

least effective.

Al Qadi et al. [22] presented a scalable approach for automatically tagging Arabic-news articles using shallow learning techniques. The study focused on the creation and utilization of two extensive datasets derived from various Arabic-news portals. The first dataset comprises 90,000 single-labeled articles spanning four domains: Business, Middle East, Technology and Sports. The second, larger dataset contains over 290,000 articles with multiple tags. These datasets have been made freely available to the Arabic computational linguistics research community, facilitating further research in the field. To validate the effectiveness of the datasets, the authors implemented ten different shallow learning classifiers. Additionally, an ensemble model was developed to combine the top-performing classifiers using a majority-voting mechanism. The classifiers demonstrated strong performance on the first dataset, with accuracy ranging from 87.7% (AdaBoost) to 97.9% (SVM). Analysis of the misclassified articles highlighted the limitations of single-label classification and underscored the importance of adopting a multi-label approach for improved accuracy.

Tahseen et al. [23] proposed the use of deep-learning techniques for detecting Arabic fake news, utilizing a dataset called AraNews, which includes articles from diverse fields, such as politics, economy, culture and sports. In their study, the authors introduced a Hybrid Deep Neural Network designed to enhance detection accuracy. This network combines the strengths of Text-Convolutional Neural Networks (Text-CNNs) and Long Short-Term Memory (LSTM) architectures. Specifically, Text-CNN is employed to extract relevant features, while LSTM handles the long-term dependencies within the text sequences. The hybrid model's performance was evaluated against the individual performances of Text-CNN and LSTM architectures. The results demonstrated that the Hybrid Deep Neural Network achieved superior accuracy, with a score of 0.914, compared to 0.859 for Text-CNN and 0.878 for LSTM.

In recent years, the detection of fake news has become increasingly important due to the rapid spread of misinformation on social-media platforms. While significant research has focused on English-language fake-news detection, there has been a notable gap in addressing this issue for the Arabic language. A significant contribution to filling this gap was by developing a large and diverse Arabic fake news dataset and employing advanced transformer-based models for classification [24], which utilize eight state-of- the-art Arabic contextualized embedding models, including AraBERT and QaribBERT, to achieve a remarkable accuracy exceeding 98%. This work not only highlights the challenges inherent in Arabic natural-language processing, such as variations in dialects and the complex structure of the language, but also demonstrates the effectiveness of transformer models in tackling these challenges. The study provides a comprehensive comparison with other fake news detection systems, underscoring the robustness of these models in accurately identifying fake news in Arabic. This research serves as a foundational reference for subsequent studies aiming to improve the accuracy and reliability of fake-news detection in non-English languages, particularly in the context of Arabic.

Despite the proliferation of studies focusing on English-language data, there is a conspicuous research gap concerning the evaluation of Arabic BERT models in the context of fake-news detection. To addresses this gap, rigorously evaluating and comparing the performance of various Arabic BERT models have been carried out by using the recently published CT23-dataset [24], which comprises a diverse array of Arabic tweets. This research is pivotal in advancing the understanding of how Arabic BERT models can be effectively leveraged to combat misinformation in the Arabic language. The study thoroughly assesses the performance of five prominent models—"Arabic Base BERT," "AraBERTV2.0," "CamelBERT MSA," "ArBERT," and "MarBERT"—revealing that "AraBERTV2.0" outperforms the others with a remarkable accuracy rate of 96%. These findings not only highlight the potential of Arabic BERT models in addressing the challenge of fake-news detection, but also provide valuable insights into the disparities among existing models, offering pathways to further enhance the precision of fake-news detection in Arabic. This work contributes significantly to the broader discourse on fake-news detection and underscores the necessity of continuous evaluation and adaptation of detection methods to keep pace with the evolving tactics of misinformation.

3. HYBRID APPROACH ARCHITECTURE

Designing an effective classifier entails a series of crucial steps, including text pre-processing and model training using deep-learning algorithms. The pre-processing of texts involves employing various approaches from the field of natural language processing (NLP), while the training phase focuses on fitting the model to pre-processed texts. In this study, a combination of NLP techniques was utilized to clean the datasets and extract the features. Subsequently, Convolutional Neural Networks (CNNs) [25], Gated Recurrent Units (GRUs) [26] and Long Short-Term Memory (LSTM) [26] were chosen for training the model.

3.1 Approaches and Proposed Model

The architecture of our hybrid approach is depicted in Figure 1, focusing on the development of a categorization system capable of classifying unclassified texts. A pre-processed sentence serves as the input to the model. Subsequently, the word vector is obtained through the process of word embedding. The model comprises several key components, including the pre-processing phase and the training phase, each of which will be detailed in the following paragraphs.

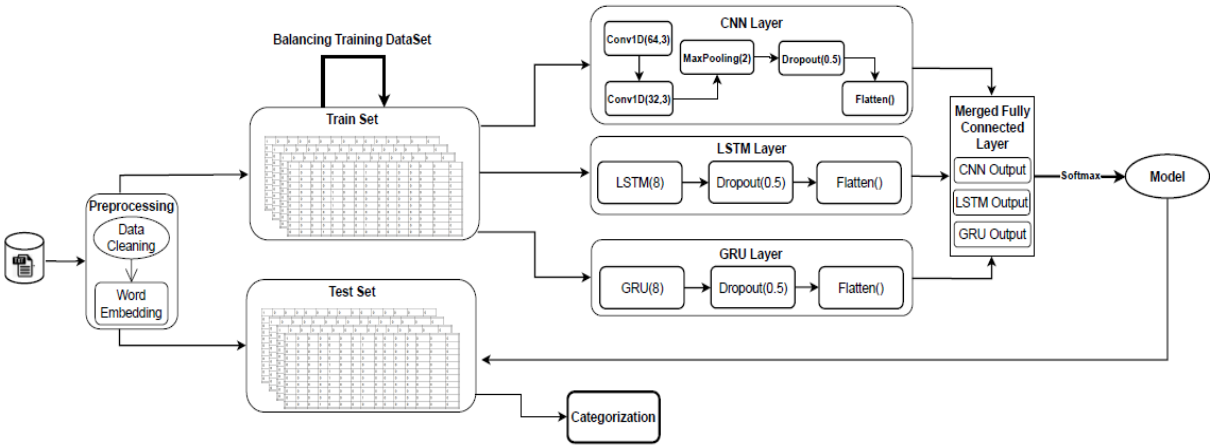


Figure 1. The hybrid architecture of the proposed multi-channel model CLGNet.

3.2 Derivation Patterns and Pre-processing

In Arabic, the majority of terms are derived from roots using standard patterns, enabling the derivation of various word forms, including verbs, adjectives, nouns and adverbs. The pattern associated with a word determines its properties, such as number (plural | singular), gender (feminine | masculine) and tense (present, past and future) [27].

Given the challenges of utilizing Arabic text directly in the training step to construct the model, a pre-processing step becomes necessary prior to training. This pre-processing step comprises two main components: text cleaning and feature extraction.

- **Text Cleaning**

The cleaning process of our datasets initiates with the removal of diacritics (tashkeel) from the text, followed by eliminating non-Arabic letters or numerals. Stemming normalization is then applied to replace various versions of the same word with its normalized or root form. Next, stop words (common words devoid of useful information) are removed, along with any unnecessary letters or symbols. Finally, the text undergoes tokenization, wherein it is segmented into its constituent words and phrases.

- **Feature Extraction**

Feature extraction aims to extract the most pertinent information from a dataset and represent it in a machine-readable format. One-hot encoding [28] is a prevalent method for describing categorical variables as binary vectors. In machine learning, categorical data like words or integers, is often transformed into a numerical representation suitable for input into machine-learning models. One-hot encoding entails generating a new binary feature for each distinct category within a feature. For

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 10, No. 04, December 2024.

example, if a feature has three distinct categories (A, B and C), three additional binary features—one for each category—are created. The original feature is then replaced by these three new binary features, with the value corresponding to the original category set to 1 and all other values set to 0.

3.3 Training and Model Selection

Following the completion of the pre-processing step, a layer of embedding was generated effectively using the pre-processed raw texts. This layer serves as input during the training phase of designing classifiers, which is further segmented into various parts. The initial step involves splitting the pre-processed data into training and test sub-sets, facilitated by the train/test split methodology.

A common strategy for evaluating the efficacy of a machine-learning model is the train/test split. Here, the model is trained using both the training set and the test set. The training set aids in determining the correlations between inputs and outputs, while the test set evaluates the model's ability to generalize to new inputs. In this research, balancing the datasets was crucial to enhancing the model's efficiency and performance. To achieve this, a percentage of 50% of the datasets was allocated for training and 25% for validation, with the remaining 25% being reserved for testing the classifier's effectiveness.

Balancing the dataset is vital in deep learning to prevent the model from becoming overly specialized, a condition known as overfitting. Overfitting can occur when there is a disproportionate representation of one type of data over the other. Common practices to address this imbalance include over-sampling the minority group and under-sampling the majority group, thereby redistributing the data in the training set to enhance the model's ability to generalize to new data.

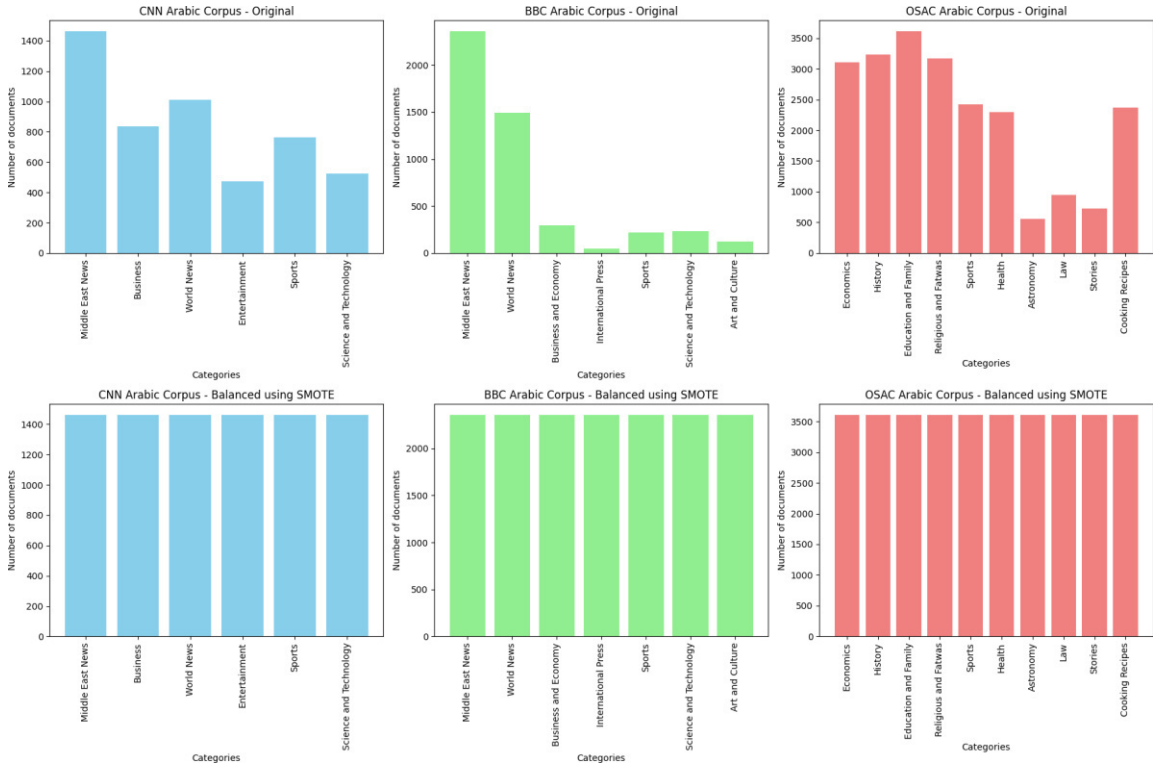


Figure 2. Class distribution of the datasets before and after balancing.

This paper utilizes SMOTE (Synthetic Minority Over-sampling Technique) to address class imbalance within our datasets. SMOTE is an effective strategy for generating synthetic samples of the minority class to match the number of data points in the dominant class. By creating new, plausible data points between existing samples, SMOTE enhances the representation of the minority class, which helps correct biases that often arise in machine-learning models trained on imbalanced datasets.

The significance of employing SMOTE lies in its ability to improve model performance and fairness. By balancing the dataset, SMOTE ensures that the model learns effectively about the minority class, leading to more accurate and reliable predictions. This is particularly important in applications where

the minority class is crucial, but underrepresented, such as medical diagnosis or fraud detection, ensuring that all classes are treated equitably and improving the overall decision-making accuracy.

Figure 2 illustrates the class distribution of three of the datasets that we used: the CNN Arabic corpus, the BBC Arabic corpus and the OSAC Arabic corpus. The figure reveals the class imbalance inherent in each dataset.

After dividing and evenly distributing the datasets, the data was input into three distinct deep-learning models: Convolutional Neural Networks (CNNs), Long Short-Term Memory networks (LSTMs) and Gated Recurrent Units (GRUs). The CNN model utilized two convolutional layers with 64 and 32 filters, respectively and a kernel size of 3. This setup allowed the CNN to detect hierarchical features within the text. Following the convolutional layers, a MaxPooling layer with a pool size of 2 was applied to down- sample and reduce dimensionality, while a dropout layer with a 0.5 probability was used to mitigate overfitting. The output from these layers was then flattened into a one-dimensional vector.

The LSTM and GRU models, each with a single layer of 8 units, were designed to handle sequential dependencies in the text. Both models included a dropout layer with a 0.5 probability after their respective recurrent layers to prevent overfitting. The outputs of the LSTM and GRU layers were also flattened into one-dimensional vectors, preparing them for integration with the CNN features.

The final step involved merging the outputs from all three models. The flattened feature vectors from the CNN, LSTM and GRU models were concatenated into a single, comprehensive feature vector. This integrated feature vector was then used to generate predictions through a final prediction model, often involving additional dense layers and a softmax activation function to output class probabilities. This hybrid approach allowed the model to leverage both local features and sequential patterns, enhancing its overall predictive performance.

Our model is mathematically expressed as follows:

Let's assume that the input to the model is a sequence of word embedding $x = x_1, x_2, \ldots, x_t$

3.3.1 CNNs

The output of the convolution operation is given in Equation 1:

$$h = relu(W_c x + b_c) \tag{1}$$

The max-pooling operation can be represented as shown in Equation 2:

$$h_{pool} = Max(h) \tag{2}$$

Finally, the results of the fully connected layer can be illustrated as in Equation 3:

$$y_{cnn} = W_f h + b_f \tag{3}$$

where, W_c is the weight matrix for the convolution operation, b_c is the bias term, W_t is the weight matrix for the fully connected layer and b_f is the bias term.

3.3.2 LSTM

At each time step t, the LSTM updates its hidden state h_t and its memory cell C_t. The hidden state h_t is used to make a prediction, while the memory cell C_t is used to retain information over time, as described in Equation 4.

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i)\ f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f)\ o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \tag{4}$$

where W_i, U_i, b_i, W_f, U_f, b_f, W_o, U_o ,and b_o are the parameters of the LSTM, x_t is the input at time step t, h_{t-1} is the hidden state at time step t 1 and σ is the sigmoid activation function. And i_t is the update gate, f_t is the forget gate, and o_t is the output gate.

$$C_t = f_t C_{t-1} + i_t \tanh(W_c x_t + U_c h_{t-1} + b_c) \tag{5}$$

where C_{t-1} is the memory cell at time step *t-1* and *tanh* is the hyperbolic tangent activation function as represented in Equation 5.

The hidden state h_t is updated as shown in Equation 6.

$$h_t = o_t \tanh(C_t) \tag{6}$$

Finally, the output passing through a fully connected layer can be represented as in Equation 7:

$$y_{lstm} = W_y h_t + b_y \tag{7}$$

where W_y and b_y are the parameters of the fully connected layer.

3.3.3 GRU

At each time step t, the GRU computes the updated hidden state h_t as shown in Equation 8:

$$\begin{aligned} z_t &= \sigma(W_z[h_{t-1} + x_t\ ] + bz) \\ rt &= \sigma(Wr[ht-1 + xt\ ] + br \\ h'_t &= tanh(W_h[r_t h_{t-1} + x_t\ ] + bh) \\ h_t &= (1 - z_t)h_{t-1} + z_t h'_t \end{aligned} \tag{8}$$

where W_z, W_r, W_h, b_z, b_r and b_h are learnable parameters of the GRU. σ is the sigmoid function used to produce a value between 0 and 1, while the tanh function maps its input to the range (-1, 1).

Finally, the final hidden state ht can be used to make predictions by passing it through a fully connected layer, as shown in Equation 9.

$$y_{gru} = W_y h_t + b_y \tag{9}$$

where W_y and b_y are learnable parameters and y is the predicted probability distribution over the classes. The final output of our model is presented in Equation 10, obtained from Equations 3, 7 and 9:

$$Y = [y_{cnn}, y_{lstm}, y_{gru}] \tag{10}$$

4. EXPERIMENTAL RESULTS AND DISCUSSION

This paragraph will delineate the achieved results both before and after balancing our datasets, followed by a discussion based on the proposed approach. Additionally, it will entail a comparison between our proposed model and other models utilizing the same datasets. We employed 50% of the datasets for training purposes, 25% for validation and reserved the remaining 25% for testing. Furthermore, our model was trained using an embedding layer size of 250, with a batch size of 512 and 10 epochs and compiled with an Adam optimizer. All experiments were conducted on a personal computer with the following specifications: an Intel(R) Core (TM) i7-8650U processor, 32 GB RAM and a frequency of 2.11 GHz.

The following datasets were utilized in both the training and testing phases of developing our model.

4.1 Datasets

To implement and train the models in this research, it is necessary to collect Arabic-news documents to use them for these purposes. These datasets were chosen due to their diversity, relevance and availability, allowing for comprehensive experimentation and evaluation of the proposed models. By leveraging these datasets, this research aims to develop a robust Arabic-news classification system that can effectively categorize news articles across different domains and topics. This research utilized three popular datasets:

4.1.1 CNN Arabic Corpus

CNN Arabic corpus is collected from CNN Arabic website cnnarabic.com. The corpus includes 5,070 text documents. Each text document belongs to 1 of 6 categories (Business 836, Entertainments 474, Middle East News 1462, Science and Technology 526, Sports 762 and World News 1010). The corpus contains 2,241,348 (2.2M) words and 144,460 district keywords after stop words removal [29].

4.1.2 BBC Arabic Corpus

BBC Arabic corpus is collected from BBC Arabic website bbcarabic.com. The corpus includes 4,763 text documents. Each text document belongs to 1 of 7 categories (Middle East News 2356, World

News 1489, Business and Economy 296, Sports 219, International Press 49, Science and Technology 232 and Art and Culture 122). The corpus contains 1,860,786 (1.8M) words and 106,733 district keywords after stop-word removal [29].

Table 1. CNN Arabic corpus dataset.

Categories	Number of documents
Middle East	1462
Business	836
World	1010
Entertainment	474
Sport	762
SciTech	526
All	5070

Table 2. BBC Arabic corpus dataset.

Categories	Number of documents
Middle East	2356
Business & Economy	296
World	1489
International Press	49
Sport	219
Science & Technology	232
Art & Culture	122
All	4763

4.1.3 OSAC Arabic Corpus

OSAC Arabic corpus is collected from multiple websites. The corpus includes 22,429 text documents. Each text document belongs to 1 of 10 categories (Economics, History, Education and Family, Religious and Fatwas, Sports, Health, Astronomy, Law, Stories and Cooking Recipes). The corpus contains about 18,183,511 (18M) words and 449,600 district keywords after stop words removal [29].

Table 3. OSAC Arabic corpus dataset.

Categories	Number of documents
Economics	3102
History	3233
Education & Family	3608
Religious and Fatwas	3171
Sports	2419
Health	2296
Astronomy	557
Low	944
Stories	726
Cooking Recipes	2373
All	22,429

4.2 Evaluation Metrics

The performance of a model can be effectively assessed by applying evaluation metrics to estimate its capabilities and construct the optimal model based on these criteria. During the research for this work, Accuracy, Precision, Recall and F1-score were employed [30].

In essence, the Accuracy metric gauges the proportion of correctly predicted cases out of the total anticipated cases. Precision is determined by dividing the number of correctly predicted positive models by the total number of models predicted as positive. Similarly, Recall measures the proportion of correctly labeled models (positively classed models). The F1-score, calculated using the harmonic mean of Recall and Precision, provides a balanced assessment of a model's performance.

Additionally, to evaluate a model's accuracy in classifying samples as either positive or negative,

statisticians commonly use the area under the receiver operating characteristic curve (AUC-ROC). Moreover, dissimilarities between expected and realized values can be quantified using techniques, such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE) or Mean Absolute Error (MAE). Lastly, the Matthews Correlation Coefficient (MCC) is utilized to evaluate accuracy while accounting for false positives.

4.3 Results and Analysis

This paper presents the findings of several experimental studies conducted to comprehensively assess the model across all dimensions. The study commenced with a comparative analysis of evaluation metrics before and after dataset balancing. Following this, overfitting graphs were generated to scrutinize the stability of the model. Subsequently, the hybrid method was tested against approaches grounded in deep-learning algorithms. Finally, a comparison was drawn with existing literature to evaluate and validate the proposed approach.

4.3.1 Comparison of Evaluation Metrics

This experiment entails comparing the outcomes achieved by our model for each dataset against those of three distinct feature-extraction approaches. The ensuing tables present the results obtained.

Table 4. Performance evaluation of the proposed model for CNN dataset.

Dataset	Accuracy	Precision	Recall	F1-score	AUC-ROC	MSE	RMSE	MAE	MCC
Balanced	94.825	94.903	95.052	94.977	0.995	0.015	0.122	0.023	0.932
Unbalanced	93.195	92.601	93.186	92.827	0.994	0.018	0.133	0.028	0.915

Table 5. Performance evaluation of the proposed model for BBC dataset.

Dataset	Accuracy	Precision	Recall	F1-score	AUC-ROC	MSE	RMSE	MAE	MCC
Balanced	90.346	90.028	79.539	83.698	0.982	0.023	0.150	0.030	0.848
Unbalanced	88.247	84.082	75.618	78.954	0.981	0.026	0.159	0.040	0.815

Table 6. Performance evaluation of the proposed model for OSAC dataset.

Dataset	Accuracy	Precision	Recall	F1-score	AUC-ROC	MSE	RMSE	MAE	MCC
Balanced	99.356	99.146	99.882	99.514	1.000	0.001	0.038	0.0024	0.989
Unbalanced	99.086	98.856	98.711	98.782	1.000	0.001	0.037	0.0026	0.989

The model's performance was assessed on both balanced and unbalanced datasets. Tables 4, 5 and 6 illustrate the results, demonstrating notably high performance, particularly for the balanced dataset. Specifically, the model achieved accuracies of 94.57%, 90.34% and 99.08% for the CNN, BBC and OSAC datasets, respectively. The precision values were 94.12%, 90.02% and 99.04%, while the recall values were 94.50%, 79.53% and 98.88% and the F1-scores were 94.30%, 83.69% and 98.93%, respectively, for the CNN, BBC and OSAC datasets.

Moreover, the model exhibited high AUC-ROC values of 0.995 for the CNN dataset, 0.982 for BBC and 1.000 for the OSAC dataset, indicating its excellent ability to distinguish between positive and negative classes. Additionally, the MSE values were 0.015, 0.026 and 0.001 for the CNN, BBC and OSAC datasets, respectively. The RMSE values were 0.122, 0.150 and 0.038 and the MAE values were 0.023, 0.030 and 0.0024, respectively. These metrics collectively suggest that the model's predictions were close to the actual class probabilities.

Furthermore, the MCC values were 0.932, 0.848 and 0.989 for the CNN, BBC and OSAC datasets, respectively, indicating the model's strong ability to correctly classify texts.

The confusion matrix in Figure 3 for the CLGNet model on the CNN dataset demonstrates its high performance across various categories. The model accurately predicts the classes with minimal errors. This high level of performance indicates the model's robustness and reliability, making it well-suited for real-world applications in news categorization and content analysis.

The confusion matrix in Figure 3 for the CLGNet model on the BBC dataset demonstrates the model's performance across various categories. The model effectively categorizes the text data, though there

are some misclassifications. This high level of performance indicates the model's robustness and highlights areas where it can be improved for even better accuracy.

The confusion matrix in Figure 3 for the CLGNet model on the OSAC dataset highlights the model's exceptional performance. This high accuracy and near-perfect precision and recall demonstrate the model's robustness in correctly categorizing the text data across diverse categories. Such performance indicates the model's potential for real-world applications in text-classification tasks, providing reliable and accurate results.

Figure 3. Confusion matrices for CLGNet using three datasets.

An error-analysis phase was conducted using the confusion matrices generated for each dataset, as shown in Table 7. These matrices provide insight into the types of errors made by the model, including false positives (FP), false negatives (FN), true positives (TP) and true negatives (TN). By analyzing the confusion matrices, we identified specific classes where the model consistently misclassified instances, particularly in cases where the classes were similar in context or had overlapping features.

Table 7. Summary of errors in confusion matrices for CNN, BBC and OSAC datasets.

Dataset	Class	True Positives	False Positives	False Negatives	Error analysis
CNN	Middle East	346	17	19	Strong classification for dominant classes
	Business	199	12	10	Confusion with World and Middle East
	World	240	13	12	Misclassification with Middle East and Business
	Entertainment	113	7	5	Struggled with classification
	Sport	181	9	9	Occasional confusion with SciTech
	SciTech	122	6	9	Some misclassification with Sport and World
BBC	Middle East	535	35	45	Consistent misclassification with World and Art
	Business & Economy	64	13	10	Confusion with Middle East and World
	World	317	39	40	Confusion with Middle East
	Int. Press	8	11	4	Some misclassification with World and Business
	Sport	46	18	14	Good classification with minor errors
	Sci & Tech	42	18	12	Some confusion with Art
	Art & Culture	20	17	7	Issues with classification, misclassified as Middle East
OSAC	Economics	774	2	1	Strong classification
	History	806	1	0	Some misclassification
	Education & Family	900	0	1	Good performance
	Religious & Fatwas	789	1	0	Strong classification
	Sports	601	0	0	Few misclassifications
	Health	573	1	0	Some misclassification with adjacent categories
	Astronomy	138	1	1	Struggled with classification
	Law	234	1	1	Issues with misclassification
	Stories	174	0	1	Some errors in classification
	Cooking Recipes	593	0	0	Strong classification, few errors

In Table 7 CNN dataset, the model showed strong performance in identifying dominant classes, like Middle East (TP: 346), but struggled with smaller, more ambiguous categories, like Entertainment, where there was a higher incidence of false negatives (e.g., 2 misclassified as Middle East and 1 as Business). Similarly, in the BBC dataset, Middle East and World were often confused due to their contextual similarities, resulting in a notable number of false positives (e.g. 28 misclassified as Middle East). The OSAC dataset presented a more diverse set of classes, with Economics (TP: 774) and

Cooking Recipes (TP: 593) being well distinguished, while smaller classes, like Astronomy and Law, showed more misclassifications (FPs and FNs). These findings suggest that the model performs well in distinguishing between major categories, but struggles with less represented or more contextually similar classes. This discrepancy suggests the need for further refinement of the model, particularly through targeted training on misclassified examples and the use of more sophisticated feature extraction methods to better differentiate between classes with overlapping features.

4.3.2 Interpretation of Training and Validation Graphs

In this experiment, we generated plots depicting the training and validation accuracy and loss values to thoroughly analyze our model's performance on the three datasets. The results are illustrated in the following graphs.

In Figures 4, 5 and 6, our model demonstrated convergence to low values in both training and validation loss curves. This convergence indicates that the model found weights that minimized the difference between predicted and actual class probabilities, as evidenced by the loss converging to near-zero values. Such convergence is crucial for effective model training and signifies that the model has learned a robust representation of the input data.

Furthermore, in addition to loss convergence, the accuracy curve also converged to values close to one. This convergence suggests that the model's representations successfully captured the fundamental patterns within the data, resulting in correct predictions. The high convergence of accuracy indicates that the model generalized well to the validation data, a significant indicator of model success.

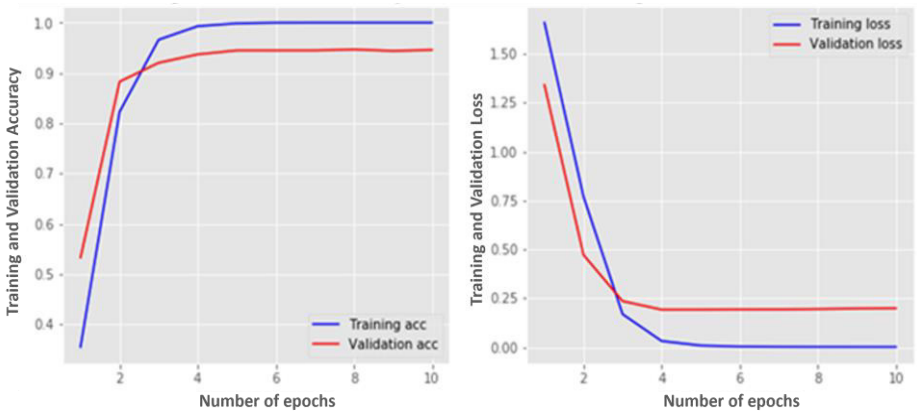


Figure 4. Graph of accuracy/loss training and validation for balanced CNN dataset.

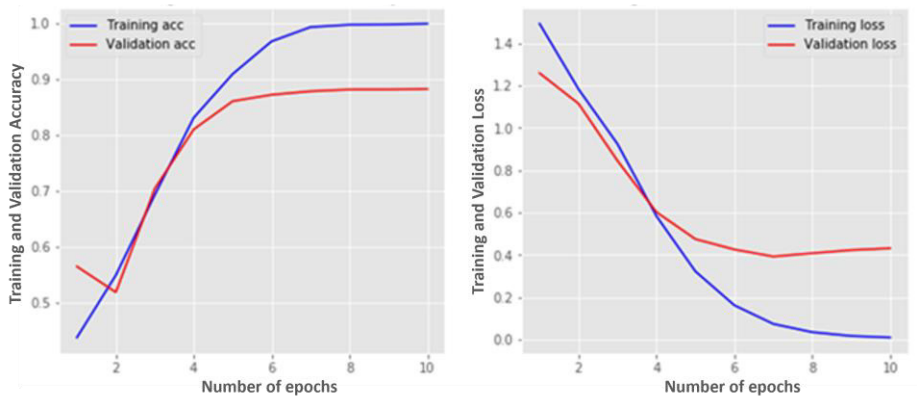


Figure 5. Graph of accuracy/loss training and validation for balanced BBC dataset.

Overall, our model demonstrated the ability to learn from training data and generalize effectively to validation data. By recognizing data patterns and generating reliable predictions, the model proves suitable for text-categorization tasks.

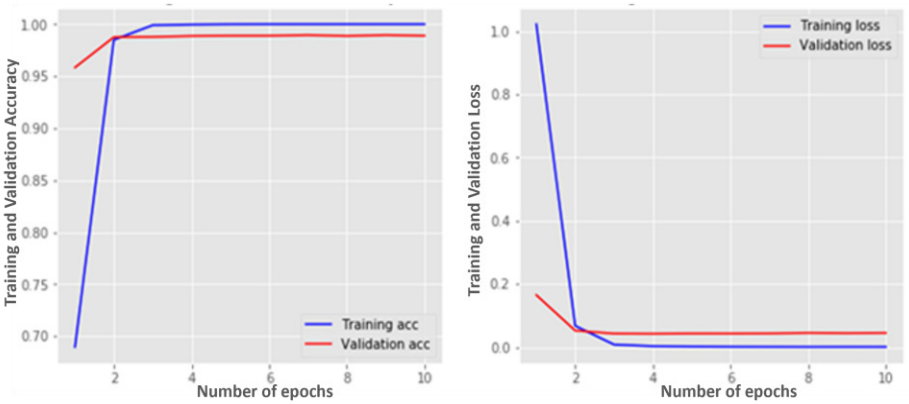


Figure 6. Graph of accuracy/loss training and validation for balanced OSAC dataset.

4.3.3 Comparison with Simple Deep-learning Layers

In this investigation, for comparative purposes, the model was segmented into three distinct components: CNN, LSTM and GRU layers. Each of these layers within our architecture was individually executed with identical parameters as the proposed model. This approach aimed to evaluate their respective performances and compare them against the performance of our proposed method.

Table 8. Comparative study of different deep-learning algorithms against the proposed model for CNN dataset.

Method	Accuracy	F1-score	Execution time (s)	Parameters
Conv1D	90.138	87.190	87.93	12,601,862
LTSM	92.110	91.443	121.24	12,532,294
GRU	93.589	92.878	95.58	12,530,246
CLGNet	94.825	94.977	669.45	12,664,390

Table 9. Comparative study of different deep-learning algorithms against the proposed model for BBC dataset.

Method	Accuracy	F1-score	Execution time (s)	Parameters
Conv1D	90.138	87.190	87.93	12,601,862
LTSM	92.110	91.443	121.24	12,532,294
GRU	93.589	92.878	95.58	12,530,246
CLGNet	94.825	94.977	669.45	12,664,390

Table 10. Comparative study of different deep-learning algorithms against the proposed model for OSAC dataset.

Method	Accuracy	F1-score	Execution time (s)	Parameters
Conv1D	90.138	87.190	87.93	12,601,862
LTSM	92.110	91.443	121.24	12,532,294
GRU	93.589	92.878	95.58	12,530,246
CLGNet	94.825	94.977	669.45	12,664,390

Tables 8, 9 and 10 reveal that the proposed CLGNet model demonstrates superior accuracy and F1-score performance compared to other methods. Despite having longer execution times and larger numbers of parameters, the performance of our proposed CLGNet model remains acceptable.

In summary, the findings suggest that CLGNet outperforms other methods in terms of accuracy and F1-score, even though it may have longer execution times and a larger number of parameters compared to GRU.

4.3.4 Comparative Study with Literature

To evaluate the proposed approach, a comparison with the literature is carried out in this context, highlighting diverse methods and results in Arabic-text classification. [31] explored the use of the ImpCHI method combined with an SVM classifier on a dataset of 5,070 Arabic documents, demonstrating superior performance with an F-measure of 90.50% when 900 features were selected. On the other hand, [32] compared traditional vectorization methods, like BOW and TF-IDF, employing classifiers, such as RL, SVM and ANN, to provide a comprehensive evaluation of their effectiveness. Meanwhile, [33] proposed the ArCAR system, a novel deep-learning approach for Arabic-text recognition and classification, achieving an accuracy of 97.76% on the Alarabiya-balance dataset, showcasing the potential of deep learning in this domain. [19] also leveraged deep learning, combining CNN and RNN models with various word embeddings, reporting high performance on the OSAC dataset, thus validating the effectiveness of hybrid architectures in capturing contextual dependencies. Finally, [34] conducted an empirical study comparing five classifiers (SVM, DT, RF, KNN and LR) using different feature-vectorization methods, finding that SVM and LR consistently outperformed others, especially when feature-vectorization techniques were applied, highlighting the stability of these classifiers across different datasets. These studies collectively emphasize the evolving landscape of Arabic-text classification, where both traditional and deep-learning methods are continuously refined to enhance accuracy and contextual understanding.

Table 11 provides a comparison of the performance of various models applied to three datasets: CNN, BBC and OSAC. The performance measure used is the F1-score, a commonly used metric for evaluating the accuracy of a classification model. The table shows that the proposed model outperforms the previous models and the other compared approaches in all datasets, with the largest improvement seen in all of them.

Table 11. Arabic-text categorization on the CNN, BBC and OSAC datasets: A comparative evaluation of recent works.

Dataset	Method	F1-score
CNN	Approach [31]	90.50
	Approach [32]	93.71
	Proposed Model	94.30
BBC	Approach [33]	69.62
	Proposed Model	83.69
OSAC	Approach [19]	98.61
	Approach [34]	98.91
	Proposed Model	98.93

To further understand the performance of our proposed model, it is important to consider the unique strengths of each component model and how they contribute to the overall effectiveness of the hybrid approach. The Convolutional Neural Network (CNN) component excels at identifying local patterns and spatial hierarchies within the text, which are crucial for recognizing key phrases that strongly indicate specific news categories. This ability to capture localized features allows CNNs to effectively process and distinguish between different types of news content. On the other hand, the Long Short-Term Memory (LSTM) network is adept at retaining long-term dependencies within sequences, making it particularly valuable for understanding the broader context of sentences where meaning may be derived from distant words or phrases. This capability enhances the model's understanding of nuanced information in longer texts, which is essential for accurate classification. Complementing these strengths, the Gated Recurrent Unit (GRU) offers a streamlined approach to handling sequential data, providing a balance between maintaining performance and optimizing computational efficiency. The GRU's simplified architecture allows the model to process large volumes of text data more quickly, without sacrificing the ability to capture necessary sequential relationships. By integrating these models, the hybrid approach leverages the CNN ability to extract critical features, the LSTM's contextual understanding and the GRU's efficiency, resulting in a robust and highly accurate Arabic news classification system.

5. CONCLUSION

Arabic-news classification presents a significant challenge in natural-language processing (NLP), requiring the categorization of news articles into predefined categories, such as politics, sports and entertainment. This task is complex due to the intricate nature of the Arabic language and the vast dimensionality of text data. This study addresses these challenges by introducing a multi-channel deep learning model, CLGNet, specifically designed for Arabic-text categorization. The model effectively analyzes and categorizes Arabic-news articles by employing data cleaning, word embedding, dataset balancing and deep-learning techniques including CNNs, LSTM and GRU. On multiple benchmark datasets, including CNN, BBC and OSAC, the model achieves impressive accuracies of 94.57%, 90.34% and 99.08%, respectively, along with F1-scores of 94.30%, 83.69% and 98.93%. Experimental results confirm the effectiveness of the proposed model, showcasing high evaluation metrics, such as Accuracy, Precision, Recall, F1-score, AUC-ROC, MSE and MCC. Our model significantly outperforms other state-of-the-art techniques in analyzing Arabic-news data, demonstrating its capability to overcome text-classification challenges and offering a promising solution for various NLP applications in Arabic. Future work aims to further enhance the model by integrating new methodologies and applying them to big-data scenarios. Additionally, the model will be evaluated on a wider range of Arabic datasets to validate its robustness and applicability across diverse linguistic domains.

REFERENCES

[1] Y. Timmerman and A. Bronselaer, "Measuring Data Quality in Information Systems Research," Decision Support Systems, vol. 126, p. 113138, DOI: 10.1016/j.dss.2019.113138, Nov. 2019.

[2] C. Porlezza, "Accuracy in Journalism," Oxford Research Encyclopedia of Communication, DOI: 10.1093/acrefore/9780190228613.013.773, Oxford University Press, Mar. 2019.

[3] N. Newman, R. Fletcher, A. Schulz, S. Andi, C. T. Robertson and R. K. Nielsen, Reuters Institute Digital News Report 2021, Reuters Institute for the Study of Journalism, pp. 1-164, 10th Edn, [Online], Available:https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2021-06/Digital_News_Report_2021_FINAL.pdf, 2021.

[4] I. Ahmad, F. AlQurashi and R. Mehmood, "Machine and Deep Learning Methods with Manual and Automatic Labelling for News Classification in Bangla Language," arXiv: 2210.10903, DOI: 10.48550/arXiv.2210.10903, 2022.

[5] R. Indrakumari, T. Poongodi and K. Singh, "Introduction to Deep Learning," in Book: Advanced Deep Learning for Engineers, pp. 1–22, DOI: 10.1007/978-3-030-66519-7_1, Springer International Publishing, 2021.

[6] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser and I. Polosukhin, "Attention Is All You Need," Proc. of the 31st Conference on Neural Information Processing Systems (NIPS 2017), pp. 1-11, Long Beach, CA, USA, 2017.

[7] H. Hassan et al., "Achieving Human Parity on Automatic Chinese to English News Translation," arXiv: 1803.05567, DOI: 10.48550/arXiv.1803.05567, 2018.

[8] M. Jabrane, I. Hafidi and Y. Rochd, "An Improved Active Machine Learning Query Strategy for Entity Matching Problem," Proc. of Advances in Machine Intelligence and Computer Science Applications (ICMICSA 2022), pp. 317–327, Springer Nature Switzerland, 2023.

[9] J. Mourad, T. Hiba, R. Yassir and H. Imad, "ERABQS: Entity Resolution Based on Active Machine Learning and Balancing Query Strategy," Journal of Intelligent Information Systems, DOI: 10.1007/s10844-024-00853-0, Mar. 2024.

[10] M. Jabrane, H. Tabbaa, A. Hadri and I. Hafidi, "Enhancing Entity Resolution with a Hybrid Active Machine Learning Framework: Strategies for Optimal Learning in Sparse Datasets," Information Systems, vol. 125, p. 102410, Nov. 2024.

[11] J. Devlin, M.-W. Chang, K. Lee and K. Toutanova, "Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding," Proc. of NAACL-HLT 2019, pp. 4171–4186, Minneapolis, Minnesota, June 2 - June 7, 2019.

[12] X. Liu, P. He, W. Chen and J. Gao, "Multi-task Deep Neural Networks for Natural Language Understanding," Proc. of the 57th Annual Meeting of the Association for Computational Linguistics, pp. 4487–4496, Florence, Italy, July 28 - August 2, 2019.

[13] P. Rajpurkar, J. Zhang, K. Lopyrev and P. Liang, "Squad: 100, 000+ Questions for Machine Comprehension of Text," Proc. of the 2016 Conf. on Empirical Methods in Natural Language Processing, pp. 2383–2392, Austin, Texas, USA, 2016.

[14] G. Lample and A. Conneau, "Cross-lingual Language Model Pre-training," Proc. of the 33rd Conf. on Neural Information Processing Systems (NeurIPS 2019), pp. 1-11, Vancouver, Canada, 2019.

[15] K. M. Fouad, S. F. Sabbeh and W. Medhat, "Arabic Fake News Detection Using Deep Learning," Computers, Materials & Continua, vol. 71, no. 2, pp. 3647–3665, 2022.

[16] M. Azzeh, A. Qusef and O. Alabboushi, "Arabic Fake News Detection in Social Media Context Using Word Embeddings and Pre-trained Transformers," Arabian Journal for Science and Engineering, DOI: 10.1007/s13369-024-08959-x, Apr. 2024.

[17] M. M. Abdelsamie, S. S. Azab and H. A. Hefny, "A Comprehensive Review on Arabic Offensive Language and Hate Speech Detection on Social Media: Methods, Challenges and Solutions," Social Network Analysis and Mining, vol. 14, p. 111, DOI: 10.1007/s13278-024-01258-1, May 2024.

[18] L. Zhang, W. Jiang and Z. Zhao, "Short-text Feature Expansion and Classification Based on Non-negative Matrix Factorization," Proc. of Machine Learning for Cyber Security(ML4CS 2020), pp. 347–362, DOI: 10.1007/978-3-030-62463-7_32, Springer International Publishing, 2020.

[19] M. S. H. Ameur, R. Belkebir and A. Guessoum, "Robust Arabic Text Categorization by Combining Convolutional and Recurrent Neural Networks," ACM Transactions on Asian and Low-Resource Language Information Processing, vol. 19, no. 5, Article no. 66, July 2020.

[20] A. M. Bdeir and F. Ibrahim, "A Framework for Arabic Tweets Multi-label Classification Using Word Embedding and Neural Networks Algorithms," Proc. of the 2020 2nd Int. Conf. on Big Data Engineering (BDE' 2020), pp. 105-112, DOI: 10.1145/3404512.340452, ACM, May 2020.

[21] A. Hassanein and M. Nour, "A Proposed Model of Selecting Features for Classifying Arabic Text," Jordanian J. of Computers and Information Technology (JJCIT), vol. 5, no. 3, pp. 275-290, Dec. 2019.

[22] L. Qadi, H. Rifai, S. Obaid and A. Elnagar, "A Scalable Shallow Learning Approach for Tagging Arabic News Articles," Jordanian Journal of Computers and Information Technology (JJCIT), vol. 6, no. 3, pp. 263-280, 2020.

[23] T. A. Wotaifi and B. N. Dhannoon, "An Effective Hybrid Deep Neural Network for Arabic Fake News Detection," Baghdad Science Journal, vol. 20, no. 4, DOI: 10.21123/bsj.2023.7427, Jan. 2023.

[24] A. B. Nassif, A. Elnagar, O. Elgendy and Y. Afadar, "Arabic Fake News Detection Based on Deep Contextualized Embedding Models," Neural Computing and Applications, vol. 34, pp. 16019–16032, May 2022.

[25] R. Romero, P. Celard, J. Sorribes-Fdez, A. Seara Vieira, E. Iglesias and L. Borrajo, "MobyDeep: A Lightweight CNN Architecture to Configure Models for Text Classification," Knowledge-based Systems, vol. 257, p. 109914, DOI: 10.1016/j.knosys.2022.109914, Dec. 2022.

[26] A. Alqahtani, H. Ullah Khan, S. Alsubai, M. Sha, A. Almadhor, T. Iqbal and S. Abbas, "An Efficient Approach for Textual Data Classification Using Deep Learning," Frontiers in Computational Neuroscience, vol. 16, DOI: 10.3389/fncom.2022.992296, Sept. 2022.

[27] A. Awajan, "Arabic Text Pre-processing for the Natural Language Processing Applications," Arab Gulf Journal of Scientific Research, vol. 25, no. 4, pp. 179–189, 2007.

[28] Y. Sun et al., "Modifying the One-hot Encoding Technique Can Enhance the Adversarial Robustness of the Visual Model for Symbol Recognition," Expert Systems with Applications, vol. 250, p. 123751, DOI: 10.1016/j.eswa.2024.123751, Sept. 2024.

[29] N. Alalyani and S. Larabi, "NADA: New Arabic Dataset for Text Classification," Int. Journal of Advanced Computer Science and Applications, vol. 9, no. 9, DOI: 10.14569/IJACSA.2018.090928, 2018.

[30] M. Hossin and M. N. Sulaiman, "A Review on Evaluation Metrics for Data Classification Evaluations," Int. Journal of Data Mining & Knowledge Management Process, vol. 5, no. 2, pp. 01–11, DOI : 10.5121/ijdkp.2015.5201, 2015.

[31] S. Bahassine, A. Madani, M. Al-Sarem and M. Kissi, "Feature Selection Using an Improved Chi-square for Arabic Text Classification," Journal of King Saud University - Computer and Information Sciences, vol. 32, no. 2, pp. 225–231, Feb. 2020.

[32] I. Jamaleddyn and M. Biniz, "Contribution to Arabic Text Classification Using Machine Learning Techniques," Proc. of Business Intelligence (CBI 2021), pp. 18–32, DOI: 10.1007/978-3-030-76508-8_2, Springer, 2021.

[33] A. Y. Muaad, H. Jayappa, M. A. Al-antari and S. Lee, "ArCAR: A Novel Deep Learning Computer-aided Recognition for Character-level Arabic Text Representation and Recognition," Algorithms, vol. 14, no. 7, p. 216, DOI: 10.3390/a14070216, July 2021.

[34] T. Sabri, O. E. Beggar and M. Kissi, "Comparative Study of Arabic Text Classification Using Feature Vectorization Methods," Procedia Computer Science, vol. 198, pp. 269–275, DOI: 10.1016/j.procs.2021.12.239, 2022.

ملخص البحث:

في عصر الصّحافة الرّقمية، يمثل تصنيف الأخبار باللّغة العربية تحدّياً مهمّاً إلى الطّبيعة المعقّدة للُّغة والتّنوُّع الكبير في المحتوى.

تعرض هذه الورقة نموذجاً مبتكراً متعدّد القنوات يستند إلى التّعلُّم العميق، مصمَّماً من أجل تحسين دِقّة تصْنيف الأخبار باللّغة العربية. ويعمل النّموذج المقترح بفاعلية على معالجة وتصنيف بيانات نَصّية باللّغة العربية.

لقد تمّ إجراء تجارب مكثَّفة على مجموعةٍ من مجموعات البيانات CNN و BBC و OSAC، حيث تمكّن النّموذج المقترح من تحقيق دقّةٍ عاليةٍ ومتانةٍ معتبرةٍ، متفوّقاً بذلك على بعض الطّرق الواردة في أدبيات الموضوع.

وتؤكّد النتائج الّتي تمّ الحصول عليها فاعلية نموذجنا الهجين (CNN، و LSTM، و GRU) في معالجة التّحدّيات المرتبطة بتصْنيف النّصوص باللّغة العربية، بالإضافة إلى إمكانية استخدامه في الأنظمة الأوتوماتيكية لتصْنيف الأخبار.

JJCIT Annual List of Reviewers (2024)

Name, *Affiliation*, Country

ABAD Maider, *Universitat Oberta de Catalunya,* Spain
ABAINIA Kheireddine, *Universit 8 Mai 1945,* Algeria
ABANDAH Gheith Ali, *University of Jordan,* Jordan
ABDELAAL Hammam M., *Luxor University,* Egypt
ABDELMOATY Ahmed, *Quebec University,* Canada
ABDELREHIM E. A., *Suez Canal University,* Egypt
ABDOLHOSSEINI Morteza, *National University of Skills,* Iran
ABDOLLAHI Rohollah, *National University of Skills,* Iran
ABDULLAH Hadeel N., *University of Technology,* Iraq
ABIMANYU Achmad, *Sepuluh November Institute of Technology,* Indonesia
ABOU EL MAJD Badr, *Mohammed V University,* Morocco
ABU SHAWAR Bayan, *Al Ain University,* UAE
AGARWAL Poonam, *Jawaharlal Nehru University,* India
AHMAD Musheer, *Jamia Millia Islamia,* India
AKHTAR Muhammad Naeem, *Lahore Garrison University,* Pakistan
AL MAMOORI Saja, *University of Windsor,* Canada
ALATTAR AlSayed R., *Zagazig University,* Egypt
ALAWIDA Moatsum, *Abu Dhabi University,* UAE
ALMURAYH Abdullah, *Imam Abdulrahman Bin Faisal University,* KSA
AL-SABRI Raeed, *Central South University,* China
ARCURI Andrea, *Kristiania,* Norway
ASLANPOUR Mohammad S., *Monash University,* Australia
AWAD Dana, *Prince Sultan University,* KSA
AYESH Aladdin, *University of Aberdeen*, UK
AZZEH Mohammad, *Princess Sumaya University for Technology,* Jordan
BAGCHI Mayukh, *University of Trento,* Italy
BAHAGHIGHAT Mahdi, *Imam Khomeini International University,* Iran
BARATA João, *University of Coimbra,* Portugal
BARBOSA Jorge Luis Victria, *University of Vale do Rio dos Sinos,* Brazil
BARKA Hazem, *Ecole de Technologie Superieure,* Canada
BELHAOUARI Samir Brahim, *Hamad Bin Khalifa University,* Qatar
BENLLOCH-CABALLERO Pablo, *University of the West of Scotland,* UK
BENSALLOUA Charef Abdallah, *University Abdelhamid Ibn Badis Mostaganem,* Algeria
BLANCO-DIAZ Cristian F., *Federal University of Espirito Santo,* Brazil
BORANDA Emin, *Manisa Celay Bayar Universiy,* Turkey
BOUJELBANE Rahma, *FSEGS,* Tunisia
BOUSSELMI Khadija, *University of Savoie Mont Blanc,* France
BUTT M. A., *Samara National Research University,* Russia
CAKIR Duygu, *Bahcesehir University,* Turkey
CAMARGO Edson T., *Federal University of Technology,* Brazil
CASTRO Francesco, *University of Bari Aldo Moro,* Italy
CENGIZ Firat, *University of Liverpool,* UK
CHAKOUR Imane, *Sultan Moulay Slimane University,* Morocco
CHEN Fulong, *Anhui Normal University,,* China
CHEN Yuhan, *Nanjing University of Information Science and Technology,* China
CHIANG Jen-Shiun, *Tamkang University,* Taiwan
COLLEY Derek Andrew, *Global Technologists Ltd.,* UK
CURIAC DanielIoan, *Politehnica University of Timisoara,* Romania
DA LUZ JNIOR Silvano Herculano, *Federal University of Pernambuco,* Brazil
DAI Qi, *North China University of Science and Technology,* China
DALARSSON Mariana, *KTH Royal Institute of Technology, Stockholm,* Sweden
DHAOUADI Asma, *University of Savoie Mont Blanc,* France
DI MONDA Davide, *University of Naples Federico II,* Italy
DIWALI Arwa, *Edinburgh Napier University,* UK
EL AKKAD Nabil, *Sidi Mohamed ben Abdellah University,* Morocco
EL GHZAOUI Mohammed, *Sidi Mohamed Ben Abdellah University,* Morocco
ELHAJJ Hassan, *MPIWG-Berlin,* Germany
ERSALI Cihan, *Batman University,* Turkey
FAHEEM Muhammad, *University of Vaasa,* Finland
FARGHALY Ali, *Polygence,* USA
FAROOQ Umar, *University of Science and Technology,* Pakistan
FENG Wei, *Panzhihua University,* China
FU Yaping, *Qingdao University,* China
FU Yu, *Beijing University of Posts and Telecommunications,* China
GAO Tan, *UCAS,* China
GAO Xuehui, *Shandong University of Science and Technology,* China

JJCIT Annual List of Reviewers (2024)

Name, *Affiliation*, Country

GAOL Ford Lumban, *Binus University,* Indonesia
GHARSELLAOUI Hamza, *Manouba University,* Tunisia
GHOBADI Seyed Eghbal, *Technische Hochschule Mittelhessen Universit of Applied Science,* Germany
GUERRERO-MENDEZ Cristian D., *State University of Campinas,* Brazil
GUPTA Brij, *Asia University,* Taiwan
GVEN Aykut Fatih, *Yalova University,* Turkey
HABASH Nizar, *New York University,* UAE
HAJBI Soufiane, *University Ibn Tofail,* Morocco
HAMDI Ahmed, *University of La Rochelle,* France
HAMEED Aroosa, *Carleton University,* Canada
HAMMO Bassam H., *Princess Sumaya University for Technology,* Jordan
HAMOUMA Moumen, *University Batna 2,* Algeria
HAN Ping, *Civil Aviation University of China,* China
HARDIE Andrew, *Lancaster University,* UK
HASAN Amit, *University of Connecticut,* USA
HASANAIN Maram, *Qatar Computing Research Institute,* Qatar
HASSABALLAH M., *South Valley University,* Egypt
HASSAN Md Mehedi, *Deakin University,* Australia
HASSANAT Ahmad B., *Mutah University,* Jordan
HERNANDEZ Antonio Marin, *Universidad Veracruzana,* Mexico
HIMDI Hanen, *University of Jeddah,* KSA
HINDARTO Djarot, *University of Nasional,* Indonesia
HO Qi Heng, *University of Colorado Boulder,* USA
HOLZINGER Andreas, *University of Natural Resources and Life Sciences,* Austria
HONG Danfeng, *Chinese Academy of Sciences,* China
HOSNY Khalid M., *Zagazig University,* Egypt
HOSSAIN Sabir, *Ontario Tech University,* Canada
HUANG Yi, *Chinese Academy of Sciences,* China
HUSSAIN Abid, *Sir Syed University of Engineering and Technology,* Pakistan
HUSSAIN Ibrar, *Shaheed Benazir Bhutto University,* Pakistan
IBRAHIMOV Yusif, *University of York,* UK
ILLAHI Usman, *Gomal University,* Pakistan
IMTIAZ Hafiz, *Bangladesh Uni. of Eng. and Techn.,* Bangladesh
IQBAL Javed, *Gomal University,* Pakistan
JAAFAR Fehmi, *Quebec University,* Canada
JALIL Zakia, *International Islamic University,* Pakistan
JARYANI Farhang, *Universiti Teknologi Malaysia,* Malaysia
JEMILI Farah, *University of Sousse,* Tunisia
JEMNI Sana Khamekhem, *University of Sfax,* Tunisia
JESUS Elsa, *UNINOVA,* Portugal
JIANG Ruihong, *BUPT,* China
JIM Md. Tanvir Rahman, *Pabna University of Science and Technology,* Bangladesh
JOSHI Akanksha, *University of Leon,* Spain
JUNG Jin-Woo, *Dongguk University,* S. Korea
KAHRMAN Mesud, *SDU,* Turkey
KARIM Abdul, *Korea University,* S. Korea
KARUPPIAH Marimuthu, *Presidency University,* India
KAYA Yiğit Bekir, *Istanbul Technical University,* Turkey
KHALILIA Hadi, *University of Trento,* Italy
KHAN Abdullah Ayub, *Beanzir Bhutto Shaheed University,* Pakistan
KHAN Muhammad Attique, *HITEC University,* Pakistan
KHAN Mustaqeem, *MBZUAI,* UAE
KHARE Smith K., *AU,* India
KHEDHER Mohammed Z., *Int. Institute for Quran and Islamic Sciences Computing,* Jordan
KHELIFI Fekher, *Nantes Universit Polytech Nantes,* France
KHEZRI Edris, *Islamic Azad University,* Iran
KHRISSI Lahbib, *Sidi Mohamed ben Abdellah University,* Morocco
KIANI Farzad, *FSMV University,* Turkey
KIANI Saad Hassan, *Ilma University,* Pakistan
KIANI Saad Hassan, *Ilma University,* Pakistan
KIANI Sina, *Semnan University,* Iran
KIEHN Jesper, *EG,* Denmark
KIM Myeongsoo, *Georgia Tech,* USA
KRSTIC Dragana, *University of Nis,* Serbia
KUS Anil, *Toros University,* Turkey
LAGHARI Asif Ali, *Shenyang Normal University,* China
LAI Qiang, *East China Jiaotong University,* China
LE Linh, *Kennesaw State University,* USA
LI Jiatao, *SHMTU,* China

Name, *Affiliation*, Country

LI Lei, *SJTU,* China
LI Wenxuan, *Beijing Language and Culture University,* China
LIANG JunGe, *Jiangnan University,* China
LIM Tong Ming, *Tunku Abdul Rahman University of Management and Technology,* Malaysia
LIU Alex, *University of Washington,* USA
LIU Fangming, *HUST,* China
LIU Hongjun, *University of Jinan,* China
LIU Huipeng, *Dalian Maritime University,* China
LIU Lei, *CUHK,* China
LIU Pengbo, *Dalian Maritime University,* China
LONG Wei, *Guizhou University,* China
LUO Shengjie, *Peking University,* China
MACEDO Jos Nuno, *University of Minho,* Portugal
MAHMOOD Arif, *ITU,* Pakistan
MAIMOUNI Mustapha, *Mohammed V University,* Morocco
MAINUDDIN Md, *Florida State University,* USA
MALAKOUTI Seyed Matin, *Amirkabir University of Technology,* Iran
MALHAS Rana, *Qatar University,* Qatar
MEI Liye, *HBUT,* China
MELMAN Anna, *National Research University Higher School of Economics,* Russia
MEZGHANI Anis, *Higher Institute of Industrial Management,* Tunisia
MEZZOUDJ Saliha, *University of Algiers,* Algeria
MOALLA Imen, *University of Sousse,* Tunisia
MODIPA Thipe I., *University of Limpopo,* South Africa
MOHANTY Sachi Nandan, *Singhidunum University,* Serbia
MUJAHID Muhammad, *Jilin University,* China
MURPHY-ARTEAGA Roberto S., *INAOE,* Mexico
MUSLIM Ali M., *Universiti Putra Malaysia,* Malaysia
N. Z. Jhanjhi, *Taylors University,* Malaysia
NAJAM Rayyan, *Islamic University of Medina,* KSA
NAMASUDRA Suyel, *National Institute of Technology Agartala,* India
NI Jiacheng, *AFEU,* China
NICA Elvira, *Academy of Economic Studies from Bucharest,* Romania
NOBANEE Haitham, *Abu Dhabi University,* UAE
NOH Jiseong, *TEEware, Inc.,* Korea
NOUBIGH Zouhaira, *Sousse University,* Tunisia
ODEH Ammar Jamil, *Princess Sumaya University for Technology,* Jordan
OGWo Chukwuebuka Elozona, *Temple University,* USA
OKUR Halil Ibrahim, *Iskenderun Technical University,* Turkey
OLADIPO Stephen, *University of Johannesburg,* South Africa
OMRI Mohamed Nazih, *University of Sousse,* Tunisia
OUAMRI Mohamed Amine, *IUT de Villetaneuse,* France
PANDEY Gaurav K., *National Institute of Technology Silchar,* India
PATWA Parth, *Amazon,* USA
PAUL Liton Chandra, *Pabna University of Science and Technology,* Bangladesh
PENG Cheng, *Xidian University,* China
PREETHI P., *Manipal Institute of Technology*, India
PRIYOTI Annita Tahsin, *University of Western Ontario,* Canada
QOBBI Younes, *Mohamed I University,* Morocco
RAHUL Kumar, *NIFTEM,* India
RAJE Vaishali V., *Krishna Institute of Medical Sciences,* India
RANI Priya, *University of Galway,* Ireland
RANI Tithi, *Rajshahi University of Engineering and Technology,* Bangladesh
RASHID Kazi Sharmeen, *University of Maryland,* USA
REZAEI Pejman, *Semnan University,* Iran
RIOS-FIGUEROA Homero V., *Universidad Veracruzana,* Mexico
ROHANINEZHAD Mohammadreza, *Urmia University,* Iran
ROMERO-GONZLEZ JulioAlejandro, *Universidad Autnoma de Quertaro,* Mexico
ROUZI Mohammad Dehghan, *University of California,* USA
ROY Arunabha M., *Texas AM University,* USA
RUDNICHENKO Nicolay, *NUOP,* Ukraine
RUSTAD Supriadi, *Dian Nuswantoro University,* Indonesia
SABTY Caroline, *German International University,* Egypt
SAEED Malik Khizar, *COMSATS University,* Pakistan
SAIDI Rakia, *UTM University,* Tunisia
SAIH Mohamed, *USMS,* Morocco
SAVCI Hüseyin Serif, *Medipol University,* Turkey
SCHWARTZ Reva, *National Institute of Standards and Technology,* USA
SEAL Ayan, *PDPM IIITDM,* India
Sebastian Volker, *HS Gesundheit University of Applied Sciences,* Germany
SEO Junyoung, *Korea University,* S. Korea
SETIADI De Rosal Ignatius Moses, *Dian Nuswantoro University,* Indonesia
SHI Yuelin, *UC Berkeley,* USA
SHIRVANI Mirsaeid Hosseini, *Islamic Azad University,* Iran
SIGMAN John B., *InfiniaML,* USA
SILVA Caroline Pacheco do Espirito, *ActiveEon,* France

Name, *Affiliation*, Country

SMYTH Dylan, *Munster Technological University,* Ireland
SOBRAL Andrews, *ActiveEon,* France
SOYLU Emel, *Samsun University,* Turkey
SPOLADORE Daniele, *STIIMA,* Italy
SUPRAPTO Falahah, *Telkom University,* Indonesia
SUYYAGH Ashraf, *University of Jordan,* Jordan
TANG Shengeng, *Hefei University of Technology,* China
TANVEER Muhammad, *University of Management and Technology,* Pakistan
TAYYAB Moeen, *International Islamic University,* Pakistan
TEMURNIKAR Ankit, *Corporate Institute of science and Technology,* India
TESTI Niccolo, *University of Macerata,* Italy
Tomoaki Ishiyama, *Chiba University,* Japan
TRUICA Ciprian-Octavian, *UPB,* Romania
TSENG Hsien-Wen, *CYUT,* Taiwan
ULLAH Shahid, *Shenzhen University,* China
UMER Qasim, *Hanyang University,* S. Korea
VALES-ALONSO Javier, *UPCT,* Spain
VELAYUDHAN Divya, *Khalifa University,* UAE
VILLANUEVAA Felix Jesus, *University of CastillaLa Mancha,* Spain
VIRIRI Serestina, *University of KwaZuluNatal,* South Africa
VISUNA Lara, *Carlos III University of Madrid,* Spain
WAN Weixiang, *University of Electronic Science and Technology of China,* China
WANG Ding, *Nankai University,* China
WANG Helin, *Johns Hopkins University,* USA
WANG Wencheng, *Weifang University,* China
WANG Wenchuan, *NCWU,* China
WANG Wenxiang, *Gannan University of Science and Technology,* China
WANG Xianzhi, *University of Technology Sydney,* Australia
WANG Zitong, *Tongji University,* China
WAQAS Muhammad, *National University of Computer and Emerging Sciences*, Pakistan
WEI Yao, *University of Twente,* Netherlands
WEI Zhe, *Shenyang University of Technology,* China
WU Jiajie, *Hangzhou Dianzi University,* China
WU Jinsong, *University of Chile,* Chile
XIE Xi, *University of Connecticut,* USA
XU Mingdi, *Wuhan University,* China
XU Qiang, *University of Huddersfield,* UK
XU Rongbin, *Putian University*, China
XU Yang, *Harbin Institute of Technology,* China
XUEJX Junxiao, *Zhejiang Lab,* China
YADAV Devendra Kumar, *XIM University,* India
YAN Mingfu, *Chinese Academy of Sciences,* China
YANG Dandi, *UC3M,* Spain
YANG Hongwen, *Beijing University of Posts and Telecommunications,* China
YANG Jian, *Alibaba Damo Academy,* China
YANG Zongyuan, *Beijing University of Posts and Telecommunications,* China
YE Zhaoyi, *HBUT,* China
YEDIDA Rahul, *North Carolina State University,* USA
YI Weichao, *Beijing Institute of Technology,* China
YILMAZ Yildiran, *Recep Tayyip Erdogan University,* UK
YIN Shoulin, *HITHarbin Institute of Technology,* China
YUAN Tao, *Shenzhen University,* China
YUE Yaoting, *FUDAN,* China
ZEINELDEEN Mohammad, *AppTek RWTH Aachen University,* Germany
ZHANG Degan, *TJUT,* China
ZHANG Jiang-bo, *Sun Yat-sen University,* China
ZHANG Shuai, *University of Stuttgart,* Germany
ZHANG Ting, *Tsinghua University,* China
ZHANG Xuan, *Johns Hopkins University,* USA
ZHANG Zhongwei, *Henan University of Technology,* China
ZHENG Jinhua, *Xiangtan University,* China
ZHOU Munyaradzi, *Midlands State University,* Zimbabwe
ZIVKOVIC Miodrag, *Singidunum University,* Serbia
ZRIBI Ines, *University of Sfax,* Tunisia
ZTRK Emir, *Trakya University,* Turkey

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

الموقع الإلكتروني: www.jjcit.org

البريد الإلكتروني: jjcit@psut.edu.jo

العنوان: جامعة الاميرة سمية للتكنولوجيا، شارع خليل السكاكت، الجبيهة، عمان، الأردن.

صندوق بريد: 1438 عمان 11941 الأردن

هاتف: 962-6-5359949+

فاكس: 962-6-7295534+

جامعة الأميرة سميّة للتكنولوجيا
Princess Sumaya University for Technology

صندوق دعم البحث العلمي والابتكار
Scientific Research and Innovation Support Fund

المجلة الأردنية للحاسوب وتكنولوجيا المعلومات

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

كانون الأول ٢٠٢٤ المجلد ١٠ العدد ٤

JJCIT

www.jjcit.org

jjcit@psut.edu.jo

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