WILEY

Research Article

An Intelligent COVID-19-Related Arabic Text Detection Framework Based on Transfer Learning Using Context Representation

Abdullah Y. Muaad [iD],[1] **Shaina Raza,**[2] **Md Belal Bin Heyat** [iD],[3] **Amerah Alabrah** [iD],[4] **and Hanumanthappa J.** [iD][1]

[1]*Department of Studies in Computer Science, University of Mysore, Manasagangothri, Mysore 570006, India*
[2]*Toronto Metropolitan University, Toronto, Canada*
[3]*CenBRAIN Neurotech Center of Excellence, School of Engineering, Westlake University, Hangzhou, Zhejiang, China*
[4]*Department of Information Systems, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia*

Correspondence should be addressed to Abdullah Y. Muaad; abdullahmuaad9@gmail.com, Md Belal Bin Heyat; drmbbh@szu.edu.cn, Amerah Alabrah; aalobrah@ksu.edu.sa, and Hanumanthappa J.; hanumsbe@gmail.com

The misleading information during the coronavirus disease 2019 (COVID-19) pandemic's peak time is very sensitive and harmful in our community. Analyzing and detecting COVID-19 information on social media are a crucial task. Early detection of COVID-19 information is very helpful and minimizes the risk of psychological security which leads to inconvenience in daily life. In this paper, a deep ensemble transfer learning framework with an understanding of the context of Arabic text COVID-19 information is proposed. This framework is inspired to spontaneously analyze and recognize the text about COVID-19. The ArCOVID-19Vac dataset has been used to train and test our proposed model. A comprehensive experimental study for each scenario is performed. For the binary classification scenario, the proposed framework records better evaluation results with 83.0%, 84.0%, 83.0%, and 84.0% in terms of accuracy, precision, recall, and $F1$-score, respectively. For the second scenario (three classes), the overall performance is recorded with an accuracy of 82.0%, precision of 80.0%, recall of 82.0%, and $F1$-score of 80.0%, respectively. In the last scenario with ten classes, the best evaluation performance results are recorded with an accuracy of 67.0%, a precision of 58.0%, a recall of 67.0%, and $F1$-score of 59.0%, respectively. In addition, we have applied an ensemble transfer learning model for this scenario to get 64.0%, 66.0%, 66.0%, and 65.0% in terms of accuracy, precision, recall, and $F1$-score, respectively. The results show that the proposed model through transfer learning provides better results for Arabic text than all state-of-the-art methods.

1. Introduction

Social media is one of the most popular communication ways which allow people to share and discuss their opinions [1]. The use of social media platforms became more widespread, providing massive opportunities for people to connect. Twitter is one of the most reliable sources of information [2, 3]. Twitter on other platforms has become a superior resource for many COVID-19 rumors and misleading information. These platforms provide incredible communication between individuals and institutions, but unfortunately, they open the door to the misuse of these social media to spread hate speech, rumors, and misleading information [4].

In Wuhan in China, a new disease called COVID-19 emerged in December 2019 [5, 6]. COVID-19 is one of the world's fastest-spreading epidemics, affecting nearly every country on the planet. The World Health Organization

declared COVID-19 a health disaster in 2020. It is the most prevalent disease in the last three years, which has confronted humanity in many countries [7]. One of the most prominent issues is that many users freely communicate through social media where they comment and post their opinions and thoughts.

The main problem is that outbreak of the COVID-19 pandemic has led to an overwhelming amount of information being shared online, making it challenging to distinguish between reliable and misleading information [8]. This study proposes an intelligent COVID-19 information detection framework that leverages a deep transfer ensemble learning model for an effective understanding of Arabic text representation (ATR). So, many electronic offenses may happen and affect people and make them in difficult psychological conditions, especially with the spread of COVID-19. In addition, monitoring of the Internet by the organization or government is weak with an increased number of users. So, this situation became an information epidemic, and it is essential to address this problem and find a solution to discover and detect this information and stop these phenomena [9].

The computer science community has taken care of these challenges by catching harmful comments and encountering them in all available ways via artificial intelligence (AI) [10, 11] since defending society is an important job that must be considered. Simultaneously, the spread of misinformation news was more that caused many victims to lose their effort and money and even exceeded their mental health. On the other side, there is an increase in the number of Arabic texts on social media. In addition, the Arabic language has a large number of dialects. Moreover, other characteristics of Arabic text such as ambiguity and being morphological of these reasons make Arabic text detection (ATD) more difficult than in other languages such as English. Thus, controlling and detecting these harmful tweets such as fake news, misinformation, and so on have become a necessity for governments, society, and individuals. Therefore, the main question is, how can a deep transfer ensemble learning model be effectively utilized to improve the understanding of ATR in the context of COVID-19 information detection? For this, it is required to address the main objective starting by developing and designing a new model which can accurately detect Arabic text while avoiding various well-known problems.

Many researchers have proposed several models for detecting COVID-19 information from social media. For example, an ensemble technique for detecting and tracking COVID-19 rumors has been used [12]. The authors [13] proposed a DL-based model. The authors [14] addressed detecting counter fake news about COVID-19 in Arabic tweets. In [9], the authors investigated DL models to help in studying COVID-19 for society's attitude. From other side, many researchers studied how COVID-19 affected our life in many scenarios using different models such as dual-level representation [15], advanced deep neural networks [16], multimodal fusion [17], and multiscale feature extraction with fusion [18]. Compared to the previous work, the main contributions of this research are summarized as follows:

(i) Develop a model to detect COVID-19 information in binary classification scenarios (i.e., non-informative data vs. informative data) as the first scenario and classify the opinion of users about the vaccine (i.e., positive, negative, and natural information) for the second scenario. Finally, detecting and classifying the document into the right class among ten classes.

(ii) The implementation model uses several AI-based techniques, i.e., ten machine learning (ML) and deep ensemble transfer learning classifiers.

(iii) The proposed DL detection framework for identifying COVID-19 Arabic tweet information is assessed using an Arabic dataset called ArCOVID-19Vac [19].

(iv) A comprehensive evaluation process is performed to select the optimal DL classifier for higher performance and fast detection of the information about COVID-19 Arabic text on social media.

(v) Compare the effectiveness of DL and ML on the ArCOVID-19Vac dataset for different scenarios by implementing a deep transfer learning model.

2. Related Works

Recently, investigating the issue of rumors has become significantly important to improve society's overall national security, especially in the light of COVID-19. Next is a rundown of the most important research on the subject, with a focus on the Arabic language. Hadj Ameur and Aliane [20] introduced a system for multilabel Arabic COVID-19 fake news and hate speech detection. Their work had been assessed on 10,828 Arabic tweets which included 10 classes. They used it to train and evaluate different classification models and declared the obtained results. The system is utilized for many applications such as the detection of hate speech and many other ATD tasks. There is an emerging demand for annotated datasets that tackle these kinds of problems in the context of COVID-19. Therefore, the authors of built and released AraCOVID-19-SSD1 which is a manually annotated Arabic COVID-19 sarcasm and sentiment detection dataset with 5,162 tweets [21]. To confirm the practical utility of the built dataset, it has been carefully analyzed and tested using several classification models. Alshalan et al. [13] conducted an analysis of hate speech in Twitter data in the Arabic region using DL and topic modeling. They aimed to identify hate speech related to the COVID-19 pandemic, which was posted by Twitter users in the Arabic region and to discover the main issues discussed in tweets containing hate speech.

Haouari et al. [22] introduced ArCOV19-Rumors, an Arabic COVID-19 Twitter dataset for misinformation detection composed of tweets containing claims from 27th January till the end of April 2020. They aim to support two classes of misinformation detection problems over Twitter: verifying free-text claims and verifying claims expressed in tweets. However, the limitation of this is being annotated by only one annotator. Jafarian et al. [23] aimed to draw

a comparison of the public's reaction to Twitter among the countries of West Asia (a.k.a., the Middle East) and North Africa to make an understanding of their responses regarding the same global threat. They mention that the results of this study can help improve treatment measures, macro decisions, social support, and a better understanding of people's behavior and reactions during an epidemic. The arCOV-19 dataset was presented in [24], which is the first Arabic Twitter dataset about the novel coronavirus (COVID-19) that includes propagation networks of a large subset of tweets.

Detecting inauthentic news about COVID-19 in Arabic tweets was addressed by Mahlous and Al-Laith. They collected nearly 7 million Arabic tweets about the COVID-19 epidemic using current hashtags at the time of the epidemic [14]. Khanday et al. [24] proposed a hybrid model for detecting COVID-19-related rumors. They concatenated LSTM and parallel CNN, so their model outperforms other methods. The authors of [25] provided an automatically annotated, bilingual (Arabic/English) COVID-19 Twitter dataset (COVID-19-FAKES). In [9], the authors investigated DL models to assist in studying COVID-19 for society's attitude. They operated on a DWLF technique to assign more weight to the loss function for the samples of the minority classes. At the same time, they created a new dataset called SenAIT, by merging the common emotions of the SenWave dataset with AIT datasets. Recently, the authors Qasem et al. [12] proposed a new approach based on ensemble techniques for detecting and tracking COVID-19 rumors.

Many works related to COVID-19 with different topics. The authors [15] proposed dual representation for image-text retrieval by innovative block-level and instance-level representation enhancement modules, respectively. The experimentations have used two datasets (i.e., Flickr30K and MSCOCO) that verify the superiority of their proposed model. The authors [18] used multiscale feature extraction and fusion methods in the image feature characterization and text information representation sections of the VQA system, respectively, to improve its accuracy. The authors [25] studied the prevalence and factors of anxiety during the coronavirus disease 2019 (COVID-19). They involved 88611 teachers from three cities. The overall prevalence of anxiety was 13.67%. They found that prevalence was higher for women than men and they used this information for decision-makers. The authors in [16] presented a new model called the deep neural networks-based logical and activity learning model (DNN-LALM) for enhancing thinking skills via logical and activity learning. The DNN-LALM employs sophisticated machine learning methodologies to offer tailored instruction and assessment tracking and enhanced proficiency in cognitive and task-oriented activities. Finally, Mubarak et al. [19] collected a dataset called ArCOVID-19 Vac for Arabic text which was manually annotated. These data have been studied on three topics called informativeness, fine-grained categorization (multiclass), and stance detection with accuracy equal to 86.4, 75.4, and 82.2, respectively. On the other hand, comparing COVID-19 ATD to image detection is still rare, as we can see from the numerous approaches employed to recognize images. Table 1 lists the latest studies regarding COVID-19 ATD.

3. Materials and Methods

This section explores the background information required to comprehend the remainder of this study, such as problem definition, effects of COVID-19 detection, representation, and classification for binary and multiclass problems that we employed to verify our experimentation. We have derived the proposed approaches that are as follows:

(i) We clearly define the problem that we need to be addressed. The problem is spreading the misinformation such as rumors through social media platforms especially with COVID-19 and our goal how to use ML and DL to address this problem by understanding the objectives and constraints associated with the problem.

(ii) We have conducted a study of existing works related to Arabic. We understand what has already been done, what methods have been used, and what gaps exist in the current knowledge.

(iii) We generate ideas and potential solutions through brainstorming sessions with authors. We started by preparing relevant data.

(iv) Method selection: we have chosen the most appropriate ML methods to implement in the proposed approach. We consider factors such as feasibility, resource requirements, and the ability to address the problem effectively.

(v) We have proposed and designed our model (architecture and parameters) using ML and DL.

(vi) We compare the results and performance of the proposed model with existing methods.

3.1. Proposed Architecture Model. Figure 1 shows the architecture of the proposed model for detecting and classifying Arabic text using ML. In addition, Figure 2 illustrates our proposed ensemble transfer learning model. There are four main stages, namely preprocessing, text representation (feature engineering), text detection, and evaluation of the proposed model.

The preparation text for further processing has to achieve in the preprocessing stage. After that, the procedures of ATR are then featuring extraction followed by feature selection and classification. Finally, both binary and two multiclass categorizations are performed using different ML and DL.

Figure 2 explores the network structure of the proposed ensemble transfer model. Initially, input text comes from the training dataset (input) with their label. In the first stage, the preparation text through the preprocessing stage has to be done. Then, the input text is represented at the word level using a feature extraction technique, the term frequency-inverse document frequency (TF-IDF) for ML and word embedding and transfer learning based on the context for DL. After FE has been done, data need to pass to the classification algorithm for learning patterns and finishing the classification task. The training text will pass with

TABLE 1: A summary of the latest studies related to COVID-19 Arabic text detection (ATD).

| Ref	Objective	Representation methods	Dataset	Evaluation metrics	Pros	Cons
[20]	Created a dataset called "AraCOVID19-MFH" for multilabel	BERT	AraCOVID19MFH dataset	$F1$-sarcastic with data augmentation 86 but without was 71	Good performance	Data not available
[21]	Built and released AraCOVID-19-SSD1	DL	Dataset with 5,162 tweets	All	Sarcasm and sentiment detection	Data not available
[13]	Aimed to identify hate speech related to the COVID-19 pandemic	DL and topic modeling	Twitter data	All	Twitter data in the Arabic region	—
[22]	Introduced ArCOV19-Rumors, an Arabic COVID-19 Twitter dataset for misinformation detection	—	Create COVID-19 Twitter dataset	All	Available	They do not consider different topics
[14]	Detecting inauthentic news about COVID-19 in Arabic tweets was addressed	—	Collected nearly 7 million Arabic tweets	—	—	Data not available
[7]	Proposed a hybrid model for detecting COVID-19-	Concatenate LSTM and parallel CNN	COVID-19 Twitter dataset	All		Need more time
[12]	Proposed a new approach based on ensemble techniques for	Ensemble techniques	COVID-19 Twitter dataset	All	Detecting and tracking COVID-19 rumors	Need more time
[19]	Collected data on three topics	ML and DL	A dataset called ArCOVID-19 Vac	Accuracy equal to 86.4, 75.4, and 82.2	Manually annotated	Their data more imbalance
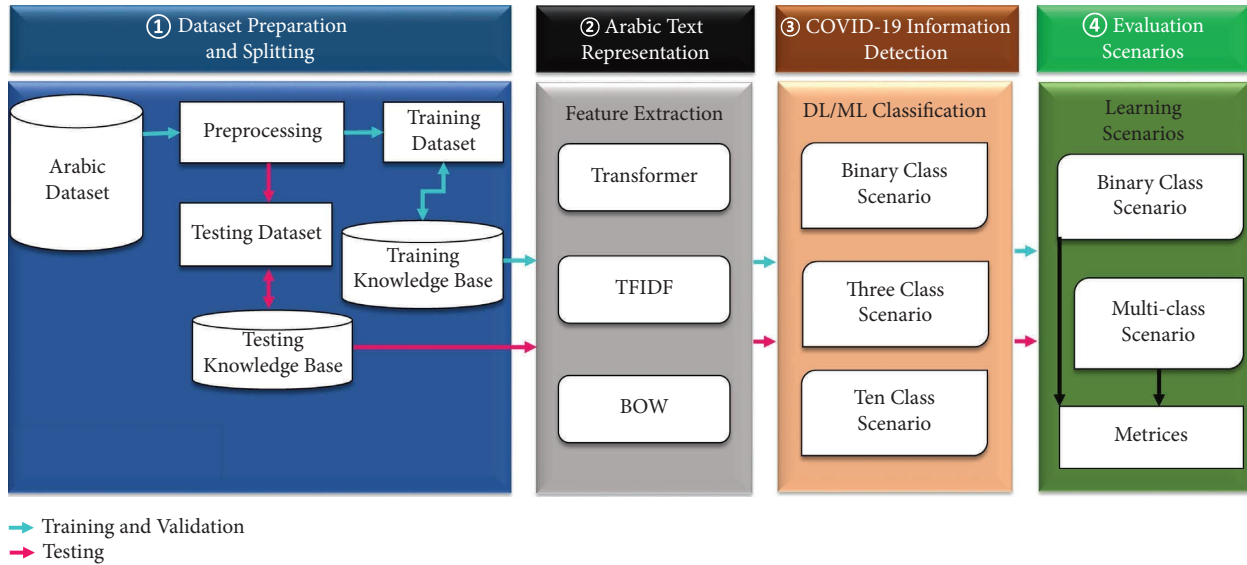
FIGURE 1: Architecture of the proposed model for detecting and classifying Arabic text using ML.
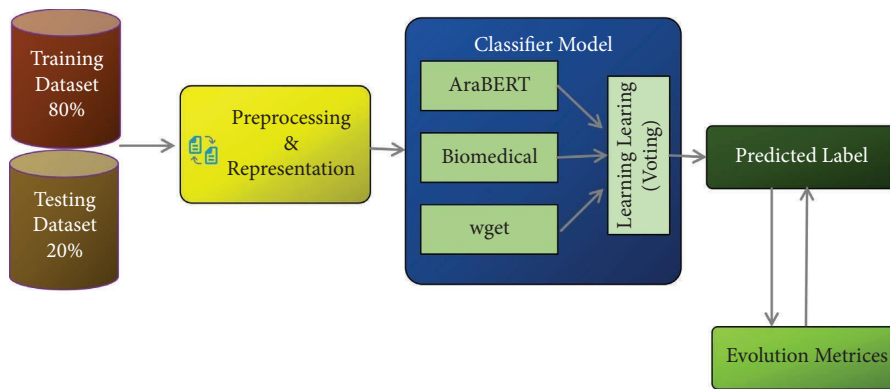

FIGURE 2: Flow chart of the proposed ensemble transfer learning model.

a corresponding label. The testing data will pass with the same process without passing the label, which will evaluate our model to predict the label and compare the predicted label with an actual label to evaluate the performance metrics of our proposed model.

3.2. Preprocessing. Preprocessing the data and preparing it for representation to learn the pattern is the first step. Preprocessing is converting data into a format that can be used easily by ML algorithms to process effectively. Some of the preprocessing steps [26] are

(i) Tokenization: this is the process of breaking the input text into individual words (or tokens). This is usually done by splitting the text based on spaces or punctuation.

(ii) Removal of non-Arabic words: this involves scanning each token and removing it if it does not conform to the Arabic script.

(iii) Stop word removal: this involves removing common words that are usually ignored by search engines and other applications, such as "and," "the," and "is."

(iv) Stemming: this is the process of reducing inflected words to their word stem, base, or root form. Arabic stemming can be complicated due to the rich morphology of the Arabic language and is typically handled by specialized Arabic NLP libraries.

3.3. Arabic Text Representation (ATR). After data preprocessing, the data need to be represented in a way that machine learning algorithms can process. This is typically done by converting the text into vectors. There are two main steps in ATR:

(i) Feature extraction: this is the step where the textual information is transformed into a set of features (numerical values). Bag of Words (BoW), which

involves representing the text as a 'bag' (set) of its words, is one such approach that is used. The text is represented as a vector where each dimension corresponds to a specific word in the BoW and the value represents the frequency of that word in the text.

(ii) AraBERT representation: the downside of the BoW model is that it does not consider the order of the words and their semantic relationship with each other. AraBERT [27], a variant of BERT specifically designed for Arabic, solves this problem. The entire text is represented as a sequence of these vectors. AraBERT uses a transformer-based architecture to model the contextual relationships among words.

3.4. Arabic Text Detection (ATD) and Classification.

Classification aims to understand the main text and classify it into the right class/category. AraBERT, a variant of BERT, has been proposed for Arabic text in 2020 [27]. We used the AraBERT model for text classification. It can be utilized for contextualized representation for different tasks, such as text understating examples, text classification, and text generating such as text translation and text summarization.

The vectors produced in the previous representation stage are then used as input to the classification algorithms. For instance, a binary classification function in a model can be represented mathematically as follows:

$$F: R^n \rightarrow \{1, 2, \ldots, k\}, \tag{1}$$

where R^n is the n-dimensional real vector space (input features) and $\{1, 2, \ldots, k\}$ is the set of the target classes. The classification function can be defined as

$$F(x) = 1/\left(1 + e^{-x^T}\beta\right), \tag{2}$$

where $x \in R^n$ is the input vector, $\beta \in R^n$ is the set of parameters to be learned, and $x^T\beta$ is the dot product of x and β. However, each classification algorithm, namely, ensemble gradient boosting classifier (EGBC), logistic regression classifier (LRC), random forest classifier (RFC), linear SVC classifier (LSVC), decision tree classifier (DTC), K-nearest neighbors' classifier (KNNC), ensemble bagging classifier (EBC), passive-aggressive classifier (PAC), and extra tree classifier (ETC), will have its own mathematical formulation and way of learning the parameters of its classification function.

3.5. Ensemble Learning.

This technique combines multiple learning models to improve overall performance. The idea is to train several classifiers and combine their predictions in some way (majority voting, weighted voting, etc.) [28, 29]. The ensemble model can be represented as

$$E(x) = G(F1(x), F2(x), \ldots, Fm(x)), \tag{3}$$

where each $Fi(x)$ is a base classifier, m is the number of classifiers, and G is the function that combines the outputs of the base classifiers. For example, in majority voting, G can be defined as

$$G(y1, y2, \ldots, ym) = \text{argmax}_i \sum \left[y_j = i\right], \tag{4}$$

where $[y_j = i]$ is the indicator function, equal to 1 if $y_j = i$ and 0 otherwise, and the sum is over all y_j.

This overall process can be used for COVID-19 information detection in Arabic text by training the classifiers on a relevant dataset. The classifiers can learn to distinguish between different types of information based on the patterns in the AraBERT representation of the text.

3.6. Implementation Environment.

The experiments were conducted on Colab Notebook and with different Python ML libraries and GPU environments. To execute the code, ML libraries, such as sci-kit-learn (https://scikit-learn.org/stable/), Keras (https://keras.io/), and TensorFlow (https://www.tensorflow.org/), have been used to finalize this model, and for fine-tuning AraBERT, we use the huggingface transformer library (https://huggingface.co/docs/transformers/index). These algorithms are deployed for different COVID-19 detection tasks for Arabic text. The datasets (https://alt.qcri.org/resources/ArCovidVac.zip) and codes are available on GitHub (https://github.com/abdullahmuaad9).

The suggested approach, "An Intelligent COVID-19 Information Detection Framework Based on Deep Transfer Ensemble Learning Model for Understanding of ATR," offers several novel aspects that contribute to the field of ATD and COVID-19 information detection. The specific novelties of this approach can include the integration of deep transfer learning and incorporation of ensemble learning techniques, where multiple classification models are combined to improve overall performance. Further, focus on COVID-19 information detection addressing context representation which captures morphology, dialectal variations, and contextual nuances specific to AT, which contributes to more accurate and context-aware COVID-19 information detection. Overall, the novelty of this suggested approach lies in its integration of deep transfer learning, ensemble learning, focus on COVID-19 information detection, consideration of ATD, and intelligent adaptability. These aspects contribute to advancing the understanding of AT, specifically in the context of COVID-19, and provide a unique and valuable contribution to the field of Arabic text classification and information processing.

3.7. Evaluation Metrics.

In this study, we have used different metrics to evaluate our work. Accuracy, precision, recall, and $F1$-score have been used [31–34]. The mathematical definition of these matrices is in equations (5)–(8) as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + TN + FP}, \tag{5}$$

$$\text{Precision} = \frac{TN}{TN + FP}, \tag{6}$$

$$\text{Recall} = \frac{TP}{TP + FN}, \tag{7}$$

$$F1 - \text{score} = \frac{2 \times TP}{2 \times TP + FP + FN}, \tag{8}$$

where positives, true negatives, false positives, and false negatives are denoted by the letters TP, TN, FP, and FN. All these parameters have been used to derive confusion matrices for both classification scenarios: binary and multiclass problems.

4. Results and Discussion

In this section, we discuss the proposed methods for the Arabic text Covid-19 classification task. In the beginning, different models have been investigated. In this work, we performed the implementation using Colab due to all libraries, and GPUS are available. The steps of this work are illustrated in Figure 1. First, traditional ML algorithms have been executed to learn the pattern in the training phase and predict the label for test files. Secondly, a transfer learning model for the Arabic language called AraBERT has been implemented. After that, a transfer to five cross-validations has been performed to get a better result. The traditional ML and transfer learning experimentation were carried out for three scenarios, binary and multiclass with three and ten classes. In the following sections, we will explain each part in detail.

4.1. Dataset. The dataset we used to evaluate the proposed AraBERT and different ML for Arabic COVID-19 is prepared for different tasks with two, three, and ten classes. The data are split into 80% for training validation and 20% for testing. The dataset details are demonstrated in the following sections, Figure 3.

The COVID-19 dataset is prepared for different tasks as shown in Tables 2–4. Table 2 shows the data distribution in the case of binary classification scenarios (i.e., non-informative vs. informative data).

Table 3 shows the stance data distribution. In this case, three classes, positive, negative, and natural information, have been considered.

Table 4 shows the fine-grained content data distribution. In this case, ten classes have been considered: info-news, celebrity, plan, requests, rumors, advice, restrictions, personal, unrelated, and others.

4.2. COVID-19 Detection Based on Binary Classification. The investigation on binary classification has been accomplished with nine ML models to make this study more effective and compare different with different models. The performances of each model are shown in Table 5. The LRC classifier has the best results compared to other ML models in terms of accuracy, precision, recall, and *F*-1 score. Compared to ML with the DL model, AraBERT has given better results than all ML models, including the LRC. The excellence of the AraBERT classifier is because the AraBERT work considers the context level, but ML lost the semantic, syntactic, and context of the text. However, the AraBERT requires more time and memory. Thus, we observed that the

proposed model with five cross-validations was excellent. A detailed comparison of all metrics for binary classification is shown in Figures 4 and 5 for all models.

4.3. COVID-19 Detection Based on Three-Class Classification. The investigation on multiclass classification has been performed with nine ML models where we compare the obtained results. The performance of each model is shown in Table 6. The LSVC classifier is the best among ML models in terms of accuracy and *F*1- score. However, the KNNC has the highest recall, while LRC has the best precision. When comparing ML and DL, AraBERT gives better results than all ML. AraBERT considers the context level, while ML loses the semantic, syntactic, and context of the text. Thus, AraBERT requires more time and memory. We notice that the classifier of our ensemble model with five cross-validations is better than the ML classifier. All performance metrics for multiclassification with three classes have been shown in Figures 6 and 7.

4.4. COVID-19 Detection Based on Ten-Class Classification. The experiment on multiclass classification was performed with nine ML models, and we compared the obtained results. The various performance metrics for each model are shown in Table 7. The LSVC classifier was the best classifier among ML models in terms of accuracy, precision, and *F*1-score. However, regarding the recall, KNNC is the best. When comparing ML and DL, AraBERT provides better results than all ML. Figures 8 and 9 show a detailed comparison between all models regarding the achieved accuracy, precision, recall, and *F*1-score.

4.5. Extension Experimentations Using Different Evaluation Metrics. In this study, we use two existing models to study in detail and we add four models to make this work more effective. These models are called MNBC, BNBC, SGDC, and SVC. To prove how the proposed model works, we extend our work by choosing scenario number two. This scenario has three classes. The algorithm of the work is mentioned in Table 8. Our study here is to see how the accuracy is affected by preprocessing and feature selection as we see in Table 9. But based on Table 10, we come to know that accuracy is not enough to evaluate the proposed model.

4.6. Comparisons between Existing and Proposed Models. Due to the limitation of the availability of the Arabic dataset in COVID-19, so we have used a new dataset which has been published in 2022. There is very less work accomplished for this dataset, and for this reason, we have implemented nine ML models to compare with. We compare our results at the beginning with the authors who publish these data. We notice that the results after proper preprocessing and applying our proposed transfer ensemble learning have got a better result for both 2-class and 3-class scenarios; except in the third scenario, the result in our proposed model was less because the authors [19] merged and decreased the classes to 4 classes instead of 10, but in our case, we keep the
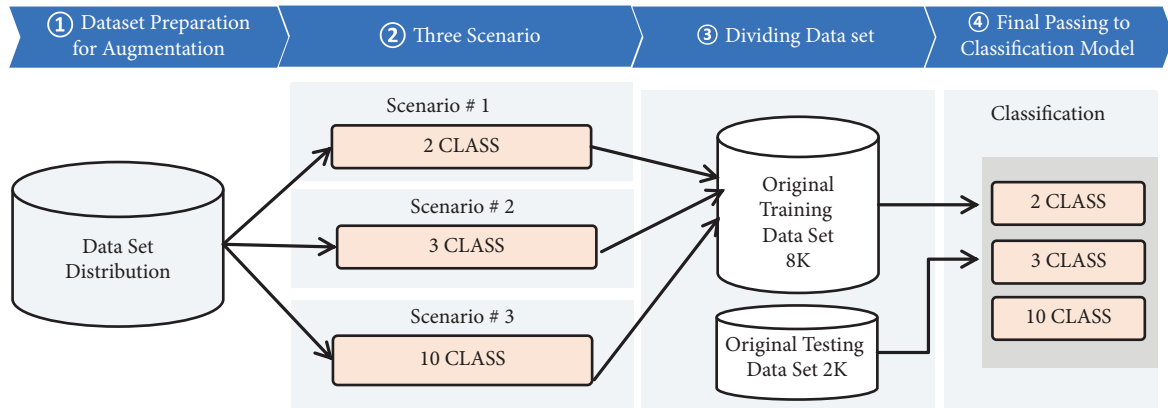
FIGURE 3: Taxonomy distribution of the dataset with different scenarios.

TABLE 2: Informative data distribution for binary class scenario.

S. no.	Category	No. of Arabic documents	Training/validation	Testing
1	More informative	5,571	4,457	1,076
2	Less informative	4,417	3,533	922
	Total	9,988	7,990	1,998

TABLE 3: Stance data distribution for multiclass scenario.

S. no.	Category	No. of Arabic documents	Training/Validation	Testing
1	Positive	7,962	6,369	1,592
2	Negative	635	508	136
3	Natural	1,391	1,113	286
	Total	9,988	7,990	1,998

TABLE 4: COVID-19 data distribution for fine-grained content of multiclass classification scenario.

S. no.	Category fine-grained content	No. of Arabic documents	Training/validation	Testing
1	Info-news	5,217	4,160	1,057
2	Celebrity	1,396	1,120	276
3	Plan	860	688	172
4	Requests	172	132	40
5	Rumors	118	94	24
6	Advice	94	77	17
7	Restrictions	24	20	4
8	Personal	1,427	1,153	274
9	Unrelated	449	359	90
10	Others	229	184	45
	Total	9988	9,986	7,987

number of classes as it is in original dataset. Our proposed models have better performance measures as compared to the existing work mentioned in Table 11. We also plan to augment the data and make these data balance in future work.

4.7. Theoretical and Practical Implications. The existing research studies work based on classical representation which do not handle the context meaning of the whole document at the same time losing semantic meaning. In addition, the same representation of sentences will be the same as long as

TABLE 5: Arabic informative data detection evaluation results for binary classification task.

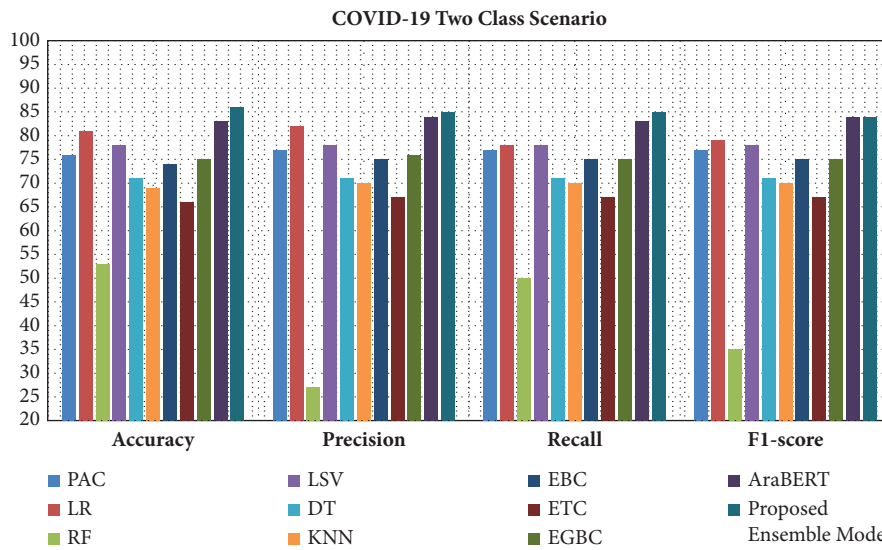| Classifier model	Accuracy	Precision	Recall	$F1$-score
PAC	0.76	0.77	0.77	0.77
LRC	0.81	0.82	0.78	0.79
RFC	0.53	0.27	0.50	0.35
LSVC	0.78	0.78	0.78	0.78
DTC	0.71	0.71	0.71	0.71
KNNC	0.69	0.70	0.70	0.70
EBC	0.74	0.75	0.75	0.75
ETC	0.66	0.67	0.67	0.67
EGBC	0.75	0.76	0.75	0.75
AraBERT	0.83	0.84	0.83	0.84
Proposed ensemble model	0.86	0.85	0.85	0.84

FIGURE 4: Comparison of the proposed model based on binary classification using accuracy, precision, recall, and $F1$-score.
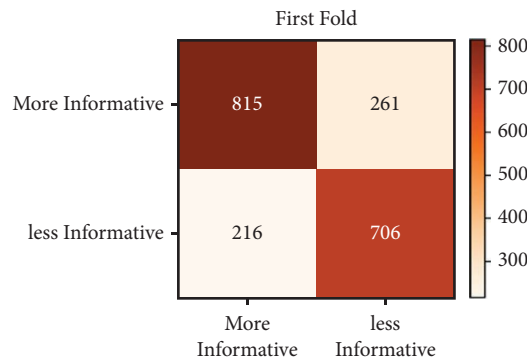

FIGURE 5: Confusion matrix of the proposed model based on binary classification.

these sentences have the same words presented in these sentences. First, the implication of this work is to study and understand the difference between classical representation and context representation for Arabic texts. The second implication is to study COVID-19 as a taxonomy problem with different scenarios starting with two, three, and ten classes. Thirdly, using ensemble learning for classification tasks adds good performance to enhance the result, but with all these advantages, the result is still not excellent due to the distribution of the dataset where there are many classes presented very less, so our model as we can see in Figure 6 do not learn the pattern for each class properly, so the future work of this study is to handle the problem of imbalance dataset issue, especially for the minority classes. So, we leave

TABLE 6: Arabic stance data detection evaluation results for ternary classification task.

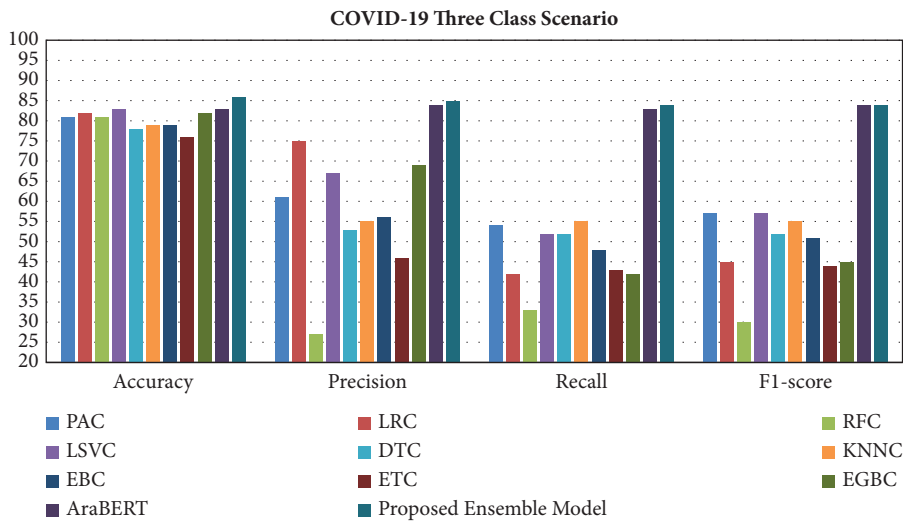| Classifier model	Accuracy	Precision	Recall	F1-score
PAC	0.81	0.61	0.54	0.57
LRC	0.82	0.75	0.42	0.45
RFC	0.81	0.27	0.33	0.30
LSVC	0.83	0.67	0.52	0.57
DTC	0.78	0.53	0.52	0.52
KNNC	0.79	0.55	0.55	0.55
EBC	0.79	0.56	0.48	0.51
ETC	0.76	0.46	0.43	0.44
EGBC	0.82	0.69	0.42	0.45
AraBERT	0.83	0.84	0.83	0.84
Proposed ensemble model	0.86	0.85	0.84	0.84

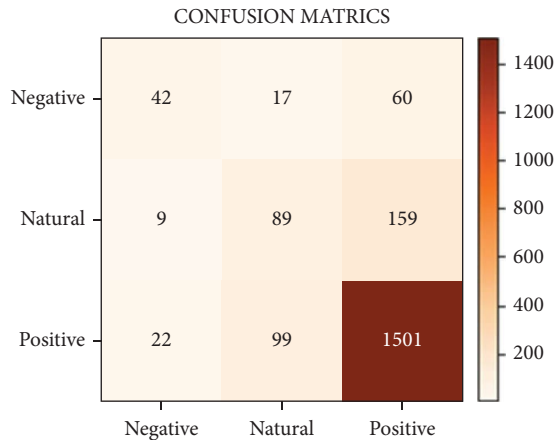FIGURE 6: Comparison of the models based on ternary classification.

FIGURE 7: Confusion matrix of the proposed model based on ternary classification.

these challenges to handle in the future with new methods such as deep active learning, few-shot learning, and augmentation techniques.

In all, the importance of ATD is critical in different scenarios such as decision-making for governments or organizations. We carefully select and optimize the proposed

TABLE 7: Arabic fine-grained content multiclass scenario data detection evaluation results for ten-class classification task.

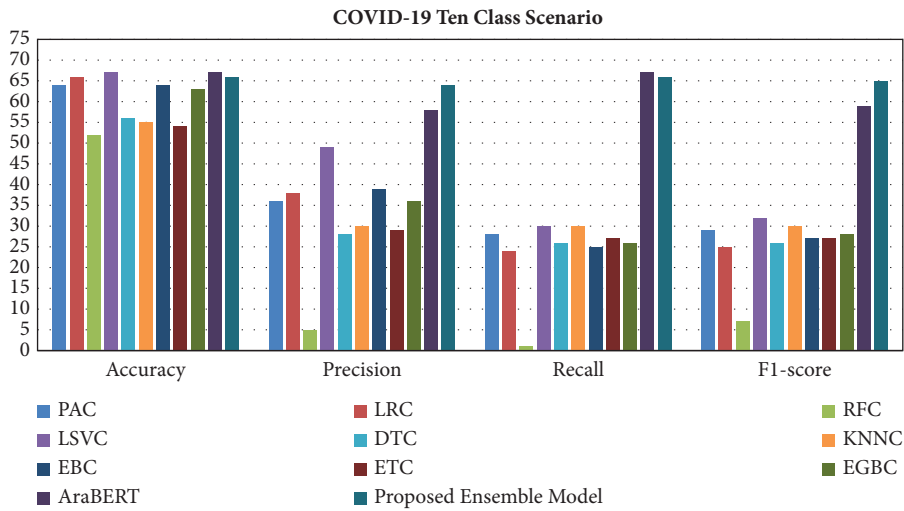| Classifier model	Accuracy	Precision	Recall	F1-score
PAC	0.64	0.36	0.28	0.29
LRC	0.66	0.38	0.24	0.25
RFC	0.52	0.05	0.01	0.07
LSVC	0.67	0.49	0.30	0.32
DTC	0.56	0.28	0.26	0.26
KNNC	0.55	0.30	0.30	0.30
EBC	0.64	0.39	0.25	0.27
ETC	0.54	0.29	0.27	0.27
EGBC	0.63	0.36	0.26	0.28
AraBERT	0.67	0.58	0.67	0.59
Proposed ensemble model	0.66	0.64	0.66	0.65

FIGURE 8: Comparison of the models based on ten-class classification.
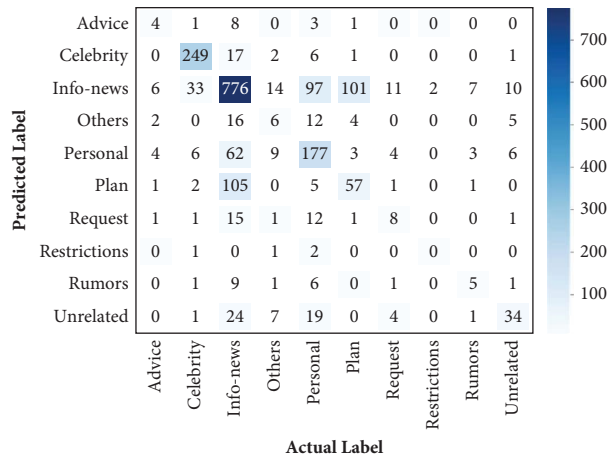

FIGURE 9: Confusion matrix of the proposed model based on ten-class classification.

TABLE 8: Pseudocode for the classification of Arabic text using (proposed model).

Inputs:	$D = \{D_1, D_2 \ldots \ldots, D_n\}$ (i) where No of Arabic text documents (ii) D_i selected document (iii) Where $\forall (D_i) \exists C_j$ (C is the name of the class) and (j) number of classes
Output:	Assign $D_{i(\text{unknownDocument})}$ to correct class C_j
Begin	Read all collections of documents in (corpus) **For** $D_{i\text{WHERE}}$ I = 1 to n Do pre-processing for document D[I]1 ← Normalization(D[i]) D[I]2 ← stopword removal (D[i]1) IF D[I]3 ← Stemming(D[i]2) IF D[I]4 ← Tokenization(D[i]3) D[I] ← TF-IDF(D[i]3) OR proposed D[I]Train = 80% D[I]Train = 20%
"Training phase"	W = Input matrix with weights train (TF-IDF/Proposed; document) Weight (W) for document and add label (L) for each document ML ← ($W + L$) where W is referred to text and L refers to label ENSEMBLE ← ($W + L$) where W is referred to text and L refers to label DL(Proposed) ← ($W + L$) where W is referred to text and L refers to label
"Testing phase"	W = Input matrix with weights train (TF-IDF/Proposed; document) ML ← ($W + L$) where W is referred to text and L refers to label ENSEMBLE ← ($W + L$) where W is referred to text and L refers to label DL (proposed) ← ($W + L$) where W is referred to text and L refers to label **End for** Push vector value without a corresponding label to classification algorithm then let the algorithm predict L Proposed ensemble transfer Learning ← (text) Class (L) ← proposed ensemble transfer learning (predict)
End	

TABLE 9: Accuracy results (%) of the ArCovidVac dataset with three classes using BoW and TF-IDF for representation with and without preprocessing.

Classifier model	BoW		TF-IDF	
	With pre	Without pre	With pre	Without pre
LRC	80	81	81	80
LSVC	79	79	80	80
MNBC	80	81	80	80
BNBC	80	80	75	73
SGDC	79	80	80	80
SVC	80	80	81	80

TABLE 10: Performance results of accuracy, recall-macro, precision-macro, F-score-macro, recall-weighted, precision-weighted, and F-score-weighted for the COVID-19 dataset (80 : 20%) without preprocessing and without feature selection (%).

Classifier model	ACC	Macro			Weighted		
		Precision	Recall	F-score	Precision	Recall	F-score
LRC	82	69	43	47	79	82	78
LSVC	82	66	53	57	80	82	80
MNBC	81	52	34	31	75	81	73
BNBC	81	52	34	31	75	81	73
SGDC	83	68	51	55	80	83	80
SVC	82	67	42	45	78	82	77

TABLE 11: Comparison between existing and proposed works.

No. of classes	Ref	Accuracy	Precision	Recall	F1-score
2	[14]	83	78	81	80
	Proposed	86	85	85	84
3	[14]	82	61	65	62
	Proposed	86	85	84	84
10	Proposed	66	64	66	65

model to be capable of handling the dialect of the Arabic language and Modern Standard Arabic (MSA) as well. This is because most of the Tweeters wrote in their special slang or dialect. In addition, we prove practically that the data representation could improve the overall performance.

4.8. Future Work. By exploring these avenues, future research in ATD can contribute to advancements in handling language variations, improving preprocessing techniques, leveraging domain adaptation and transfer learning, addressing data scarcity, exploring multilingual and cross-lingual approaches, promoting fairness and ethics, and applying text classification to real-world Arabic language applications. The limitations of the current study on Arabic text classification may include challenges of Arabic language such as orthographic, morphological, phonology, and various dialects. This article confirms that many open issues need to be addressed, including the limitation of the lack of availability of benchmark datasets and the lack of dictionaries and lexicons of Arabic texts. Moreover, there was the difficulty of the nature of the Arabic language in terms of morphology and delicate. One of the important problems is imbalanced data which need to be designed and we proposed a new data augmentation technique to get better performance; all these are suggestions to handle in future work.

5. Conclusion

The problem of ATD is a challenging task compared to other languages such as English because of different reasons. COVID-19 has become a significant issue for people since its appearance and spread in 2019. Fake news, misinformation, and many more affected negatively our nation's life. So, in this article, we implemented a model for the detection and classification of COVID-19 for Arabic text in different scenarios, which can help in making plans, helping decision-makers, avoiding rumors, etc. We have carried out this work utilizing an ArCOVID-19Vac dataset. Our proposed transfer ensemble learning model provides excellent performance (accuracy, precision, recall, and F1-score) for the three scenarios. This article confirms that many open issues need to be addressed, including the limitation of the lack of availability of benchmark datasets and the lack of dictionaries and lexicons of Arabic texts. Moreover, the difficulty of the nature of the Arabic language in terms of morphology is delicate. One of the important problems is imbalanced data, which needs to be designed, and a new data augmentation technique to get better performance; all these are suggestions to handle in future work.

Data Availability

The data used to support the findings of this study are included in this article.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Authors' Contributions

A.Y.M conducted conceptualization, methodology, software, writing of the original draft, data curation, and visualization; S.R. conducted writing, review, and editing, data curation, software, and resources; M.B.B.H. performed conceptualization, visualization, investigation, writing of the original draft, project administration, data curation, resources, and supervision; A.A. and H.J. conducted data curation, supervision, investigation, writing, review, and editing, and funding acquisition. All authors read and agreed to the publication.

Acknowledgments

References

[1] M. Alkhair, K. Meftouh, K. Smaïli, and N. Othman, "An Arabic corpus of fake news: collection, analysis and classification," *Communications in Computer and Information Science*, vol. 1108, pp. 292–302, 2019.

[2] I. Abu Farha and W. Magdy, "Multitask learning for {A}rabic offensive language and hate-speech detection," in *Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection*, pp. 86–90, Marseille, France, May 2020.

[3] H. Mulki and B. Ghanem, "Let-mi: an Arabic levantine twitter dataset for misogynistic language," in *Proceedings of the 6th Arabic Natural Language Processing Workshop*, pp. 154–163, Kyiv, Ukraine, April 2021.

[4] M. Djandji, F. Baly, W. Antoun, and H. Hajj, "Multi-task learning using AraBert for offensive language detection," in *Proceedings of the 4th Workshop on Open-Source Arabic*

Corpora and Processing Tools, with a Shared Task on Offensive Language Detection, pp. 97–101, Marseille, France, May 2020.

[5] C. C. Ukwuoma, Z. Qin, M. B. B. Heyat 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, p. 709, 2022.

[6] C. C. Ukwuoma, D. Cai, M. B. B. Heyat et al., "Deep learning framework for rapid and accurate respiratory COVID-19 prediction using chest X-ray images," *Journal of King Saud University- Computer and Information Sciences*, vol. 35, no. 7, 2023.

[7] M. Al-Sarem, A. Alsaeedi, F. Saeed, W. Boulila, and O. Ameerbakhsh, "A novel hybrid deep learning model for detecting covid-19-related rumors on social media based on lstm and concatenated parallel cnns," *Applied Sciences*, vol. 11, no. 17, p. 7940, 2021.

[8] S. Raza and B. Schwartz, "Constructing a disease database and using natural language processing to capture and standardize free text clinical information," *Scientific Reports*, vol. 13, no. 1, p. 8591, 2023.

[9] N. Alturayeif and H. Luqman, "Fine-grained sentiment analysis of Arabic covid-19 tweets using bert-based transformers and dynamically weighted loss function," *Applied Sciences*, vol. 11, no. 22, p. 10694, 2021.

[10] S. Parveen, "Interweaving artificial intelligence and biosignals in mental fatigue: unveiling dynamics and future pathways," in *Proceedings of the 2023 20th International Computer Conference on Wavelet Active Media Technology and Information Processing*, pp. 1–9, IEEE, Chengdu, China, December 2023.

[11] F. Akhtar, "Early coronary heart disease deciphered via support vector machines: insights from experiments," in *Proceedings of the 2023 20th International Computer Conference on Wavelet Active Media Technology and Information Processing*, pp. 1–7, IEEE, Chengdu, China, December 2023.

[12] S. N. Qasem, M. Al-Sarem, and F. Saeed, "An ensemble learning based approach for detecting and tracking COVID19 rumors," *Computers, Materials and Continua*, vol. 70, no. 1, pp. 1721–1747, 2021.

[13] R. Alshalan, H. Al-Khalifa, D. Alsaeed, H. Al-Baity, and S. Alshalan, "Detection of hate speech in COVID-19-related tweets in the Arab Region: deep learning and topic modeling approach," *Journal of Medical Internet Research*, vol. 22, no. 12, Article ID e22609, 2020.

[14] A. R. Mahlous and A. Al-Laith, "Fake news detection in Arabic tweets during the COVID-19 pandemic," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 6, pp. 778–788, 2021.

[15] S. Yang, Q. Li, W. Li, X. Li, and A. A. Liu, "Dual-level representation enhancement on characteristic and context for image-text retrieval," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 11, pp. 8037–8050, 2022.

[16] D. Li, K. D. Ortegas, and M. White, "Exploring the computational effects of advanced deep neural networks on logical and activity learning for enhanced thinking skills," *Systems*, vol. 11, no. 7, p. 319, 2023.

[17] S. Lu, M. Liu, L. Yin, Z. Yin, X. Liu, and W. Zheng, "The multimodal fusion in visual question answering: a review of attention mechanisms," *PeerJ Computer Science*, vol. 9, 2023.

[18] S. Lu, Y. Ding, M. Liu, Z. Yin, L. Yin, and W. Zheng, "Multiscale feature extraction and fusion of image and text in VQA," *International Journal of Computational Intelligence Systems*, vol. 16, no. 1, 2023.

[19] H. Mubarak, S. Hassan, S. A. Chowdhury, and F. Alam, "ArCovidVac: analyzing Arabic tweets about COVID-19 vaccination," in *Proceedings of the 13th Language Resources and Evaluation Conference*, pp. 3220–3230, Marseille, France, May 2022.

[20] M. S. Hadj Ameur and H. Aliane, "AraCOVID19-MFH: Arabic COVID-19 multi-label fake news and hate speech detection dataset," *Procedia Computer Science*, vol. 189, pp. 232–241, 2021.

[21] M. S. H. Ameur and H. Aliane, "AraCOVID19-SSD: Arabic COVID-19 sentiment and sarcasm detection dataset," 2021, https://arxiv.org/abs/2110.01948.

[22] F. Haouari, M. Hasanain, R. Suwaileh, and T. Elsayed, "ArCOV19-Rumors: Arabic COVID-19 twitter dataset for misinformation detection," in *Proceedings of the 6th Arabic Natural Language Processing Workshop*, Kyiv, Ukraine, April 2021.

[23] H. Jafarian, M. Mohammadi, A. Javaheri et al., "Topic discovery on farsi, English, French, and Arabic tweets related to COVID-19 using text mining techniques," *Studies in Health Technology and Informatics*, vol. 279, pp. 26–33, 2021.

[24] A. M. U. D. Khanday, S. T. Rabani, Q. R. Khan, N. Rouf, and M. Mohi Ud Din, "Machine learning based approaches for detecting COVID-19 using clinical text data," *International Journal on Information Technology*, vol. 12, no. 3, pp. 731–739, 2020.

[25] Q. Li, Y. Miao, X. Zeng, C. S. Tarimo, C. Wu, and J. Wu, "Prevalence and factors for anxiety during the coronavirus disease 2019 (COVID-19) epidemic among the teachers in China," *Journal of Affective Disorders*, vol. 277, pp. 153–158, 2020.

[26] S. R. Bashir, S. Raza, V. Kocaman, and U. Qamar, "Clinical application of detecting COVID-19 risks: a natural language processing approach," *Viruses*, vol. 14, no. 12, p. 2761, 2022.

[27] W. Antoun, F. Baly, and H. Hajj, "AraBERT: transformer-based model for Arabic language understanding," 2020, http://arxiv.org/abs/2003.00104.

[28] L. Ali, Z. He, W. Cao, H. T. Rauf, Y. Imrana, and M. B. Bin Heyat, "MMDD-ensemble: a multimodal data-–driven ensemble approach for Parkinson's disease detection," *Frontiers in Neuroscience*, vol. 15, pp. 754058–754111, 2021.

[29] C. C. Ukwuoma, "Boosting breast cancer classification from microscopic images using attention mechanism," in *Proceedings of the 2022 International Conference on Decision Aid Sciences and Applications*, pp. 258–264, IEEE, Chiangrai, Thailand, March 2022.

[30] M. B. Bin Heyat, D. Lai, K. Wu et al., "Role of oxidative stress and inflammation in insomnia sleep disorder and cardiovascular diseases: herbal antioxidants and anti-inflammatory coupled with insomnia detection using machine learning," *Current Pharmaceutical Design*, vol. 28, no. 45, pp. 3618–3636, 2022.

[31] M. J. A. Fazmiya, A. Sultana, M. B. B. Heyat et al., "Efficacy of a vaginal suppository formulation prepared with Acacia arabica (Lam.) Willd. gum and Cinnamomum camphora (L.) J. Presl. in heavy menstrual bleeding analyzed using a machine learning technique," *Frontiers in Pharmacology*, vol. 15, pp. 1–23, 2024.

[32] C. C. Ukwuoma, Z. Qin, M. Belal Bin Heyat et al., "A hybrid explainable ensemble transformer encoder for pneumonia

identification from chest X-ray images," *Journal of Advanced Research*, vol. 48, pp. 191–211, 2023.

[33] M. B. B. Heyat, F. Akhtar, F. Munir et al., "Unravelling the complexities of depression with medical intelligence: exploring the interplay of genetics, hormones, and brain function," *Complex Intell. Syst.*, 2024.

[34] Sumbul, A. Sultana, M. B. B. Heyat et al., "Efficacy and classification of Sesamum indicum linn seeds with Rosa damascena mill oil in uncomplicated pelvic inflammatory disease using machine learning," *Frontiers of Chemistry*, vol. 12, 2024.