EAI Endorsed Transactions
on Internet of Things Research Article **EAI.EU**

Enhancing Arabic E-Commerce Review Sentiment Analysis Using a hybrid Deep Learning Model and FastText word embedding

Nouri Hicham[1, *], Habbat Nassera [2], Sabri Karim[3]

[1,3]Research Laboratory on New Economy and Development (LARNED), Faculty of Legal Economic and Social Sciences AIN SEBAA, Hassan II University of Casablanca, Morocco
[2]RITM Laboratory, CED ENSEM Ecole Superieure de Technologie Hassan II University, Casablanca, Morocco

Abstract

The usage of NLP is shown in sentiment analysis (SA). SA extracts textual views. Arabic SA is challenging because of ambiguity, dialects, morphological variation, and the need for more resources available. The application of convolutional neural networks to Arabic SA has shown to be successful. Hybrid models improve single deep learning models. By layering many deep learning ensembles, earlier deep learning models should achieve higher accuracy. This research successfully predicted Arabic sentiment using CNN, LSTM, GRU, BiGRU, BiLSTM, CNN-BiGRU, CNN-GRU, CNN-LSTM, and CNN-biLSTM. Two enormous datasets, including the HARD and BRAD datasets, are used to evaluate the effectiveness of the proposed model. The findings demonstrated that the provided model could interpret the feelings conveyed in Arabic. The proposed procedure kicks off with the extraction of Arabert model features. After that, we developed and trained nine deep-learning models, including CNN, LSTM, GRU, BiGRU, BiLSTM, CNN-BiGRU, CNN-GRU, CNN-LSTM, and CNN-biLSTM. Concatenating the FastText and GLOVE as word embedding models. By a margin of 0.9112, our technique surpassed both standard forms of deep learning.

Keywords: hybrid model, Deep learning, Arabic Sentiment analysis, FastText, E-Commerce Review

Received on 25 September 2023, accepted on 09 December 2023, published on 14 December 2023

doi: 10.4108/eetiot.4601 _______________

1. Introduction

Over the previous ten years, Twitter, LinkedIn, and Facebook have each seen significant increases in user traffic and popularity. Recently, companies and other types of organizations have realized that it is beneficial to use these platforms to engage with their clients and gain further knowledge about those customers. It can be challenging to determine an individual user's overall level of contentment with a certain brand due to the enormous number of posts, users, messages, comments, and other contact forms [1]. SA is a subclass of NLP that employs sophisticated machine learning and data mining models to measure attitudes, feelings, reactions, emotional responses, and sentiments in a variety of fields, such as quality of service, market reach, pricing, and public support of government events and activities [2]. Arabic is characterized by many irregular forms and intricate rules for morpho-syntactic agreement with various linguistic variants that englobe many sets of norms for written expression. It may only be possible to create robust generalized models from Arabic text with suitable treatment and processing of the data. The Arabic Sentiment Analysis (ASA) project has significantly fewer resources than its English counterpart. These resources offer lexicons of emotions and corpora of emotions that have been tagged. The ASA has received much attention due to these difficulties [3].

Both DL and ML have the potential to automate text analysis as well as the extraction of sentiment [4], [5]. The CNN and CNN-LSTM hybrid models have significantly

Corresponding author. Email: nourihicham@ieee.org

improved sentiment analysis [6]. The CNN deep layers include pooling, convolutional, and fully connected layers, which can extract more significant and pertinent information from text data. Memory in LSTM is organized in such a way that it can memorize important text information as well as comprehend sentences. For example, [7] presented a CNN-LSTM hybrid model that divides Arabic attitudes into two basic classes by employing a word2vec model. This model categorizes Arabic attitudes. [8] suggested performing Arabic sentiment analysis. The model was constructed using LSTM as well as RNN. Increasing one's level of productivity is possible by integrating various hybrid DL models. Combining the results of several individual classifiers results in creating an ensemble classifier. Because of this, the component models can better respond to the flaws in each other's representations, increasing the classification's accuracy. In recent years, more focus has been placed on machine learning research on learning methodologies known as ensembles. In ensemble learning (EL), which focuses on learning uniform ensembles, opinion mining could be utilized more frequently. It is anticipated that heterogeneous ensembles, constructed using a variety of core classifiers and datasets, will lead to improvements in the produced ensembles. Ensemble modelling as a technique for enhancing NLP models is becoming increasingly common [9]. Ensemble classifiers incorporate the results of several individual classifiers into a single prediction model. The performance of the combined model is anticipated to be superior to that of each of the base classifiers [10]. In particular, a learning model can adjust the weights assigned to the final ensemble modelling for each NLP basis system. This helps the learning model to produce the most accurate judgments possible. On the other hand, this training-based ensemble needs to be more accurate in situations with inadequate data. The EL process is contingent on how predictions and base learners are coupled, more specifically, on whether rule-based or meta-learning approaches are applied and whether the learning process is concurrent or sequential [11], [12]. The difference between heterogeneous and homogeneous ensembles is that heterogeneous ensembles contain many classifiers, while homogeneous ensembles contain multiple instances of the same model. The variance of both heterogeneous and homogeneous ensemble basis classifiers grows as a result of the presence of a variety of approaches. Heterogeneous ensembles can outperform homogenous ones if the biases in each contribute anything positive to the overall performance [13]. Three primary ensemble learning algorithms can be used for any given dataset. These are bagging [14], boosting [15], or stacking [16].

- Therefore, we created a hybrid deep learning model for ASA by utilizing the word embedding capabilities of both FastText and Glove. This model is con-structed using the most effective possible combination of hybrid CNN-BiLSTM.

-This model is constructed using the most effective combination of hybrid CNN-BiLSTM. We offer a model that investigates the deep learning models that have already been pre-trained. These models include CNN, LSTM, GRU, BiGRU, BiLSTM, CNN-BiGRU, CNN-GRU, CNN-LSTM, and CNN-biLSTM. Word embedding techniques such as FastText and Glove were utilized to merge the results obtained from the various deep-learning databases.

2. Related works

Even though most ASA strategies are traditional machine learning (ML), [17] used Logistic Regression (LR) to classify customer reviews by utilizing TF-IDF; the ways of categorizing emotions in Arabic and dialects were very similar. [18] investigated the comments made on Saudi social media platforms using machine learning. The Support Vector Machine (SVM), Decision Tree (DT), and Naive Bayes were all beaten by the K-Nearest Neighbour (KNN) algorithm, which achieved an accuracy of 78.46%. (NB). In [19], NB, the Rocchio classifier, and SVM are utilized to evaluate the degree of subjectivity and SA present in Arabic customer reviews. For the particular ensemble method specified, the meta-learner ensemble approach LR delivered the best possible results.

DL techniques for SA [20] have been examined for their ability to deliver greater robustness and flexibility by the autonomous feature extracting the data. CNN and RNN are two instances of such television networks. Deep neural network (DNN) techniques in the Arabic dialect SA are limited. This contrasts chat-bots, remote sensing, recommendation systems, and load monitoring.

LSTMs and CNNs were utilized by the authors [19]. The embedding matrix for the words was produced by using the word2vec program. The ASTD and SemEval 2017 results showed that their model had the best performance. CNN-LSTM hybrid models using word2vec extracted features for binary categorization of Arabic viewpoints have been reported [8]. Provided for consideration as an alternative, The Main-AHS, ASTD, and Ar-Twitter datasets, in addition to other ASA datasets, were utilized. With the CNN–LSTM hybrid model, we attained an accuracy of 79.07%. The authors of the study [20] employed CNN to analyse nine different datasets, two of which were the ASTD and the LABR. Tweets and re-views are both included in the collection. Skip-Gram and CBOW word2vec were utilized to generate the word embedding matrix. CNN was also used to balance out datasets that were not evenly distributed. To conduct Arabic sentiment analysis, an LSTM-RNN hybrid model was developed [9]. This model utilized both LSTM and RNN. One of the topics covered was examining deep learning using a variety of pre-trained word embeddings. The AraSenTi-Tweet was used to perform the testing on their model.

Ensemble models can improve model inference. It is also possible to utilize hybrid models as base classifiers within an ensemble to improve hybrid techniques. Compared to base models, ensemble models perform better. An application of ensemble modelling in ASA has been carried out. The authors employed voting to improve the accuracy of Arabic sentiment analysis. Optimization was utilized with CNN-LSTM to choose the most effective LSTM and CNN for AS tweets (ASTD). Those individuals with the highest f1 scores were chosen., the emotion gleaned from Arabic tweets can be predicted using an ensemble model. Based on CNN and LSTM, this predictive model received recognition for its accuracy. They utilized the ASTD dataset to come up with the study results. The f1 score and accuracy of the ensemble model are both the greatest possible values [21]. Ghosh et al. (2023) embarked on a comprehensive study to assess water quality through predictive machine learning. Their research underscored the potential of machine learning models in effectively assessing and classifying water quality. The dataset used for this purpose included parameters like pH, dissolved oxygen, BOD, and TDS. Among the various models they employed, the Random Forest model emerged as the most accurate, achieving a commendable accuracy rate of 78.96%. In contrast, the SVM model lagged behind, registering the lowest accuracy of 68.29% [22].

Alenezi et al. (2021) developed a novel Convolutional Neural Network (CNN) integrated with a block-greedy algorithm to enhance underwater image dehazing. The method addresses colour channel attenuation and optimizes local and global pixel values. By employing a unique Markov random field, the approach refines image edges. Performance evaluations, using metrics like UCIQE and UIQM, demonstrated the superiority of this method over existing techniques, resulting in sharper, clearer, and more colourful underwater images [23].

Sharma et al. (2020) presented a comprehensive study on the impact of COVID-19 on global financial indicators, emphasizing its swift and significant disruption. The research highlighted the massive economic downturn, with global markets losing over US $6 trillion in a week in February 2020. Their multivariate analysis provided insights into the influence of containment policies on various financial metrics. The study underscores the profound effects of the pandemic on economic activities and the potential of using advanced algorithms for detection and analysis [24].

3. Methodology

In this paper, we put up a novel method for ASA that is based on the idea of ensemble learning. A hybrid CNN, LSTM, Bi-GRU, GRU, and Bi-LSTM architecture are among the deep learning classifiers that are utilized in the technique that has been proposed. After that, the proposed method integrates the results of the classifiers with those of meta-classifiers. Using this hybrid learning strategy allows us to increase overall performance while also taking advantage of the structural and functional benefits of each model. In the following paragraphs, we will continue our discussion of classification models by delving deeper into hybrid models, the word embedding model, and the base and meta classifiers.

3.1. Word Embedding

Word embedding, often known as WE, is a form of word representation that attempts to communicate the meaning of words using vectors. The meanings and contexts of words represented by vectors comparable to one another are referred to as embedded words. In SA and other NLP tasks, we are considered crucial components, and the data processing layer they provide for DL approaches serves as the foundation of the data processing chain. As a result, the goal of this section is to present numerous types of WE used in this work.

GloVe

The goal of the GloVe [22] project is to research the global presentation of the entire corpus and incorporate the terms' meanings into this investigation. When establishing the value of the real-valued vectors linked with individual words, word frequency and co-occurrence are the most important metrics that are taken into consideration. Because the gloVe is an unsupervised method, humans are only sometimes engaged in giving the collection of words meaning. The foundation for the computation is the use of the frequency of particular words and the frequency of words immediately adjacent to each word. The first thing that has to be done to finish using GloVe is compiling a list of the terms that are used as the context the most often. In the second step of the process, an X co-occurrence matrix is constructed by going through the corpus and then through each phrase. We want to express the index of frequently occurring words with the letter n and the remaining words in the corpus with the letter a. The value of Pnm indicates the likelihood that the word m will appear in the same setting as the word b in the future.

$$G_{ab} = \mathrm{X}_{ab} / X_b \qquad (1)$$

With the use of b, a, and a third word from the context k, we can calculate the likelihood of co-occurrence of T (b, a, and k) as continues to follow:

$$T_{(b,a,k)} = \frac{\mathrm{G}_{ba}}{\mathrm{G}_{ak}} \qquad (2)$$

In conclusion, the following is the formula that may be used to compute the loss function M:

$$M = \sum_{ba=1}^{v} f(X_{ba})(W_b^T k + B_n + B_m - \log X_{ba})^2 \qquad (3)$$

The objective of the training is to achieve the greatest possible reduction in the error caused by the least squares method, where f refers to the function that assigns weights. After completing the training phase, GloVe will assign a real-valued vector to each word.

FastText

A well-known natural language processing (NLP) method that efficiently identifies and represents text is the open source fastText project developed by Facebook Research. Learning word representations is not the primary objective of fastText embedding; rather, the primary focus is investigating the fundamental structure of words. Because it releases students from memorizing their representations of words that contain numerous morphemes, this works very well in languages with a large number of morphemes[12]. The likelihood of the context, which is expressed by a word t and can be adjusted by word vectors thanks to the application of a grading function F:

$$F(w_c, w_t) = \mathrm{U}_{wc}^{C} \vee wt \qquad (4)$$

V and U are taken from the output and input matrix embeddings, respectively. The following is an explanation of how the scoring function of fastText works:

$$F(w_c, w_t) = \sum_{g \in Gwc} \wedge_g^C \vee wt \qquad (5)$$

Gwc is an abbreviation for the collection of n-grams found in the word w_c, and g represents the gth n-gram in vector form. The notation Vwt denotes the vector related to the context word w_t.

3.2. Deep learning Techniques

In this section, we discussed the deep learning models, including a conventional neural network (CNN), a GRU network, a Bi-GRU network, an LSTM network, and a Bi-GRU network. Following this is a document that presents an in-depth analysis of each model.

Convolutional neural network (CNN)

CNN [1] is typically used for computer vision, but in recent years, it has been broadened to challenges involving NLP and has produced remarkable results in Arabic. Our opinion target extraction model was upgraded with the addition of CNNs to obtain character-level attributes such as suffixes and prefixes [30]. Character vectors that have been trained using CNN. The matrix C is created by stacking the character vectors that have been searched. After that, various convolution filters with varying widths are inserted between matrix C and many filter matrices. This is done so that the pooling of character-level data can yield the best possible results. Before implementing character embedding, we used CNN's dropout layer to prevent overfitting and ensure that the model was appropriately sized.

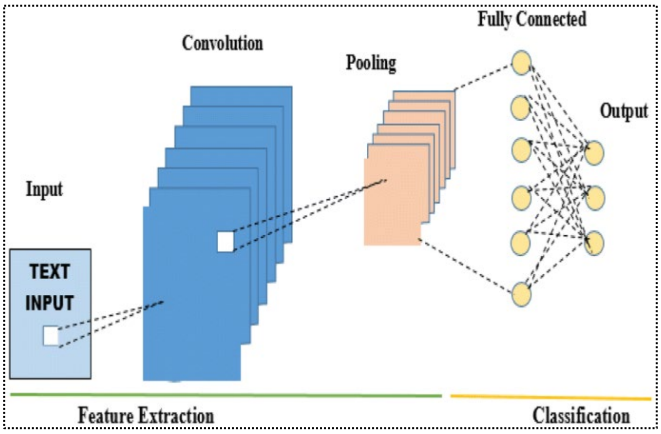

Figure 1. CNN framework for sentiment analysis

Long Short-Term Memory (LSTM)

Long Short-Term Memory is a form of recurrent neural network [23]. In RNN, the output of the previous step is used as input in the step that is now being processed. It did this by addressing the issue of the RNN's long-term dependencies, in which the RNN cannot forecast the word that is kept in the long-term memory but can make more accurate predictions based on more recent information. RNN does not deliver an efficient performance when the gap length is increased. LSTM has the capability, by design, to remember the information for a significant amount of time. Processing, forecasting, and categorizing outcomes based on time-series data are all possible applications for this tool.

Long Short-Term Memory (LSTM) is a Recurrent Neural Network (RNN) form that was developed to process sequential data like time - series data, voice, and text. It was named after its acronym, "Long Short-Term Memory." Because LSTM networks can learn long-term dependencies in sequential data, they are ideally suited for applications like language translation, speech recognition, and time series forecasting.

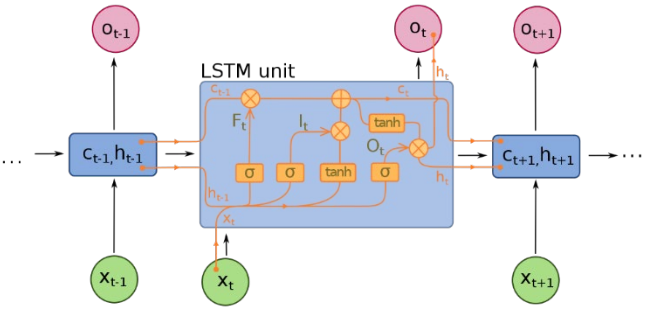

Figure 1. LSTM framework for sentiment analysis

Gated Recurrent Unit (GRU)

Since the design of both LSTM and GRU is the same, GRU can be thought of as a version of the LSTM. GRU [24] is an RNN network architecture used extensively, and since both LSTM and GRU share the same design, GRU is widely employed. The problem of disappearing gradients is solved by GRU through the utilization of update and reset gates. In addition, the update gate enables the model to determine the quantity of information that should be collected before.

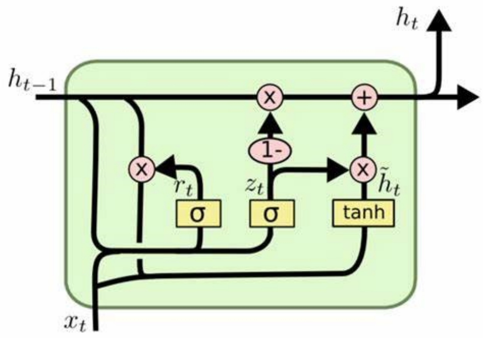

Figure 2. GRU architecture

Bi-GRU

The loss of information is a direct result of the inability of GRU networks to utilize either the present or the future environment. Many researchers have utilized a bidirectional GRU (Bi-GRU) tool, which enables them to process data in both forward and backward orientations. In this layer, the information from the hidden levels is brought up to the output level. The formation of a bidirectional GRU network begins with the combination of two GRUs. While the input sequence of one network is displayed in the conventional time order, the sequence of another network is shown in the reverse order. Each stage consists of the merged outputs of both networks [23]. The context is provided by this structure, as indicated in figure 3.

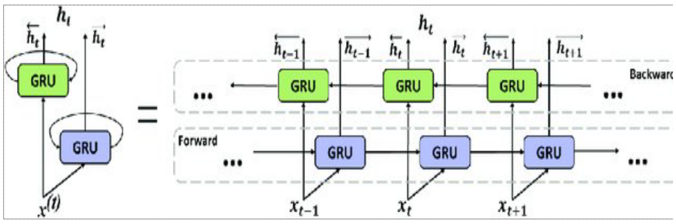

Figure 3. Bi-GRU architecture

Bidirectional LSTM (Bi-LSTM)

BiLSTM [24] is a subclass that falls under the umbrella of the more general LSTM network. Figure 4 shows that it uses a bidirectional network, which denotes that the inputs will flow in two directions: from the future to the past and from the past to the future. Additionally, the outputs will flow in the opposite direction, from the past to the future. Because of this, the network can gather information in both the forward and the reverse directions. Therefore, it can keep knowledge pertinent to the past and the future at any given time by exploiting the two hidden states and storing information about the future in memory.

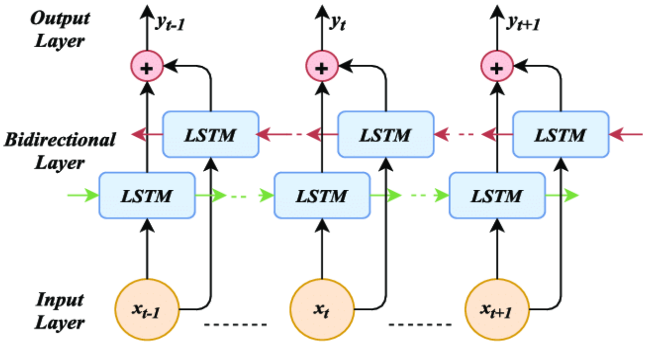

Figure 4. Bi-LSTM architecture

Hybrid deep learning model (CNN-BiLSTM)

In deep learning, hybrid models are models that mix different kinds of neural network architectures or components to take advantage of each one's strengths and make more powerful and effective models. These models often combine different types of layers or networks, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), or transformer networks, to handle specific problems or jobs.

Hybrid models aim to take advantage of how the traits and abilities of different neural network architectures work well together. Each architecture has its strengths and flaws. Combining them allows us to make models that can handle a wider range of data patterns and show complex relationships.

In our work, hybrid models combining CNNs and BiLSTMs have performed better at several NLP jobs. The CNN component helps catch local patterns and features, while the BiLSTM component captures the data's overall context and sequential dependencies. This makes for better representations and better performance.

4. Experiments and results

In the following paragraph, we will begin by discussing the assessment metric that we used in order to evaluate the performance of our model, and then we will move on to the conclusions that we acquired. When that is finished, we will then talk about our concluding remarks.

4.1. Dataset description

The initial stage in doing a sentiment analysis is the collection of textual data. The following datasets were examined in order to test the efficacy of our model within the context of this study:

- The Hotel Arabic-Reviews Dataset (HARD) [25] is comprised of 490587 hotel comments that were obtained from the Booking website throughout June and July 2016. The reviews may be found in Modern Standard Arabic (MSA), and dialectal Arabic (DA), and reviewers have given each item a total of one to ten stars. The huge HARD data set served as the foundation for our research and contained both positive and negative comments and ratings. The review count for this dataset is 93700, and it includes an equal number of positive (46850) and negative (46850) types of reviews.

The Books Reviews in Arabic Dataset, often known as BRAD [26], comprises 510,600 book reviews. The evaluations were retrieved from the GoodReads.com website between June and July 2016. MSA and DA account for the vast majority of the study. We utilized the BRAD dataset, which contains an equal amount of positive and negative comments. Only reviews with positive and negative ratings (four and five stars) are displayed (one and two stars). The dataset has a little more than 156 thousand different reviews altogether.

4.2. Performance measures

Several distinct measures are utilized to determine the efficacy of the proposed enhancement to the performance of our model. The measurements that are employed can have an impact on how the usefulness and performance of models are monitored and compared. A concise description of the five criteria we applied to evaluate the quality of our research is provided in Table 1.

Table 1. A description of evaluation metrics for sentiment analysis

Performance Metrics	Definition	Equation
Accuracy	Accuracy can be defined as the proportion of correct predictions made relative to the total number of instances.	$\frac{Tp + Tn}{Tp + Tn + Fp + Fn}$
Precision	Precision can be defined as the proportion of correctly predicted samples in relation to the total number of samples.	$\frac{Tp}{Tp + Fp}$
F1-score	The harmonic mean of a player's scores is used to calculate the F1-score, which measures their correctness and memory.	$\frac{2 * (\text{Precision.Recall})}{(\text{Precision} + \text{Recall})}$
Recall	The percentage of expected positive samples is what is meant when one talks about the recall.	$\frac{Tp}{Tp + Fn}$
Specificity	Specificity is opposed to recall.	$\frac{Tn}{Tn + Fp}$

4.3. Experimental parameters

During our studies, we made use of, among other tested parameters, the following:

Table 1. Experimental setting parameters

Setting Parameter	Score ranges	Best value
Batch size	100. 80. 64. 32. 16. 8	32
Epoch	40. 30. 20. 15	20
Optimizer	RMSprop, Adadelta, Adagrad, Adamax, Adam.	Adam
Dropout	0,6. 0,5. 0,4. 0,2. 0,0	0,2
Activation function	softmax, relu, softplus, linear, tanh,	tanh

4.4. Experimental results

In this section, we will discuss the findings of our study, which compared the impact of several different word embeddings on the classification of Arabic sentiment analysis carried out with CNN, BiLSTM, BiGRU, GRU, LSTM, CNN-BiLSTM, CNN-LSTM, CNN-GRU, and CNN-BiGRU. Specifically, this section will focus on the CNN-BiLSTM and CNN-LSTM models. When analyzing performance, we focused on Accuracy, Specificity, Precision, and F1-score. The recall was also taken into consideration.

The results obtained for the datasets containing Arabic text are presented in Tables 3 and 4. The proposed model, which employs a hybrid deep learning technique with FastText word embedding, outperforms the basic classifiers with the greatest accuracy of 92.03%, 92.58%, and 92.27% on the HARD, BRAD, and ARD datasets, respectively. This is accomplished through the utilization of a model that uses a FastText word embedding. The fact that the model performs better than the base classifiers is evidence of this claim.

When comparing the accuracy of the trained models on the HARD dataset, Table 3 reveals that our model with FastText word embedding achieved the best accuracy 87.46%. This was followed by the GRU model with FastText word embedding, which achieved an accuracy of 87.20%. The accuracy of the CNN-BiGRU hybrid deep learning model that used the FastText word embedding was 84.54 percent.

According to the findings in Table 3, our hybrid deep learning model that included a FastText word embedding achieved the best level of accuracy, which was 91.12%. The GRU model with FastText word embedding model came in second, achieving 90.85% accuracy. The CNN model trained using the Glove word embedding fared the poorest in terms of accuracy, scoring 73.00%.

Our model achieves respectable results across all evaluation metrics; Precision, , recall ,F1-score, and Specificity , scoring0.9198, 0.7358, 0.7334, and 0.7718 on the HARD dataset and 0.9583, 0.7666, 0.7641, and 0.8041 on the BRAD dataset, respectively; figure 5 and 6 presents a global view of the performance of our hybrid deep learning model.

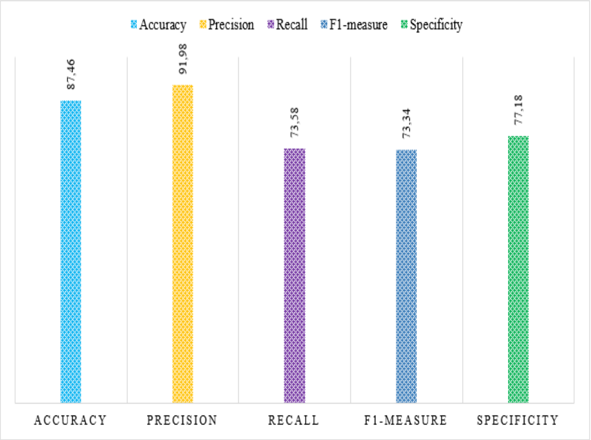

Figure 5. Performance of our model in HARD

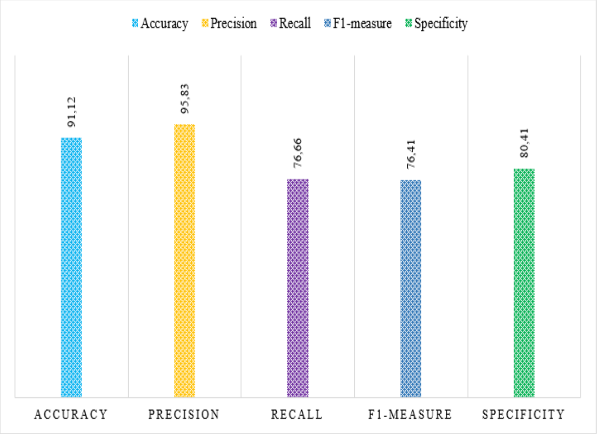

Figure 7. Performance of our model in BRAD

Table 3. Evaluation of model application in the HARD dataset

HARD dataset						
Deep learning techniques	Word Embeddings	Accuracy	Precision	Recall	F1-measure	Specificity
	Glove	0,7007	0,7525	0,4342	0,6918	0,8338
CNN	FastText	0,7526	0,8115	0,5102	0,7046	0,8514
	Glove	0,7351	0,7642	0,4343	0,6722	0,8563
BiLSTM	FastText	0,8155	0,8740	0,5033	0,7259	0,7995
	Glove	0,7761	0,8225	0,6417	0,7742	0,8338
BiGRU	FastText	0,8372	0,8401	0,6980	0,6692	0,8514
GRU	Glove	0,7367	0,7774	0,5898	0,8031	0,8845
	FastText	0,8720	0,9171	0,5343	0,7313	0,7696
LSTM	Glove	0,8145	0,7525	0,4342	0,6918	0,7836
	FastText	0,7526	0,8115	0,5102	0,6041	0,8385
CNN-LSTM	Glove	0,7784	0,8250	0,6437	0,7765	0,8362
	FastText	0,8400	0,8426	0,7000	0,6711	0,8539
CNN-BiLSTM	Glove	0,7387	0,7797	0,5915	0,8055	0,8871
	FastText	0,8746	0,9198	0,7358	0,7334	0,7718
CNN-BiGRU	Glove	0,7870	0,8402	0,7129	0,7325	0,8237
	FastText	0,8454	0,8919	0,5697	0,7357	0,6505
CNN-GRU	Glove	0,7853	0,8385	0,7112	0,7307	0,8220
	FastText	0,8438	0,8902	0,5680	0,7340	0,6488

Table 2. Evaluation of model application in the BRAD dataset

BRAD dataset						
Deep learning techniques	Word Embeddings	Accuracy	Precision	Recall	F1-measure	Specificity
	Glove	0,7300	0,7840	0,4523	0,7207	0,8687
CNN	FastText	0,7841	0,8455	0,5315	0,7341	0,8870
	Glove	0,7659	0,7962	0,4524	0,7003	0,8921
BiLSTM	FastText	0,8496	0,9106	0,5243	0,7563	0,8329
	Glove	0,8086	0,8569	0,6685	0,8066	0,8687
BiGRU	FastText	0,8722	0,8753	0,7272	0,6972	0,8870
GRU	Glove	0,7675	0,8099	0,6145	0,8367	0,9215
	FastText	0,9085	0,9555	0,5566	0,7619	0,8018
LSTM	Glove	0,8486	0,7840	0,4523	0,7207	0,8164
	FastText	0,7841	0,8455	0,5315	0,6294	0,8736
CNN-LSTM	Glove	0,8110	0,8595	0,6706	0,8090	0,8712
	FastText	0,8751	0,8779	0,7293	0,6992	0,8896
CNN-BiLSTM	Glove	0,7696	0,8123	0,6162	0,8392	0,9242
	FastText	0,9112	0,9583	0,7666	0,7641	0,8041
CNN-BiGRU	Glove	0,8199	0,8754	0,7427	0,7631	0,8582
	FastText	0,8808	0,9292	0,5935	0,7665	0,6777
CNN-GRU	Glove	0,8182	0,8736	0,7409	0,7613	0,8564
	FastText	0,8791	0,9274	0,5917	0,7647	0,6759

5. Summary and future directions

This study examines the challenge of comprehending the feelings conveyed through Arabic writing. The effectiveness of the Arabic sentiment analysis system was evaluated concerning various hybrid deep learning models based on CNN, BiGRU, BiLSTM, GRU, LSTM, CNN-BiGRU, CNN-GRU, CNN-LSTM, and CNN-biLSTM, respectively. These models were used to evaluate the system's performance.

Finding out how successful the proposed model will be required using both the HARD and BRAD datasets. The experiments' results demonstrated that the proposed model is suitable for comprehending the feelings expressed in Arabic writings. The proposed method begins with the extraction of FastText model features at the beginning of the process. After that, we construct and train deep learning models, such as CNN, CNN-BiGRU, CNN-GRU, CNN-LSTM, and CNN-biLSTM.

Confirmation of our methods came in the form of an authentic Arabic review dataset. When applied through FastText, the recommended technique beats the baseline models on the BRAD dataset, achieving an accuracy of 0.9112 and a better overall performance. The outcomes of this study shed light on the value of textual data to professionals in formulating strategies, enhancing competitiveness, and managing income. Applying the significance of the meaning of the Arabic word to subsequent initiatives will result in improved performance. In our subsequent work, we will make sure to consider this limitation.

References

[1] N. Habbat, H. Anoun, et L. Hassouni, « Sentiment Analysis and Topic Modeling on Arabic Twitter Data during Covid-19 Pandemic », Indones. J. Innov. Appl. Sci. IJIAS, vol. 2, no 1, p. 60-67, févr. 2022, doi: 10.47540/ijias.v2i1.432.

[2] M. A. El-Affendi, K. Alrajhi, et A. Hussain, « A Novel Deep Learning-Based Multilevel Parallel Attention Neural (MPAN) Model for Multidomain Arabic Sentiment Analysis », IEEE Access, vol. 9, p. 7508-7518, 2021, doi: 10.1109/ACCESS.2021.3049626.

[3] G. Badaro et al., « A Survey of Opinion Mining in Arabic: A Comprehensive System Perspective Covering Challenges and Advances in Tools, Resources, Models, Applica-tions, and Visualizations », ACM Trans. Asian Low-Resour. Lang. Inf. Process., vol. 18, no 3, p. 1-52, sept. 2019, doi: 10.1145/3295662.

[4] M. Wankhade, A. C. S. Rao, et C. Kulkarni, « A survey on sentiment analysis methods, applications, and challenges », Artif. Intell. Rev., vol. 55, no 7, p. 5731-5780, oct. 2022, doi: 10.1007/s10462-022-10144-1.

[5] N. Hicham et S. Karim, « Analysis of Unsupervised Machine Learning Techniques for an Efficient Customer Segmentation using Clustering Ensemble and Spectral Clustering », Int. J. Adv. Comput. Sci. Appl., vol. 13, no 10, p. 9, 2022, doi: 10.14569/ijacsa.2022.0131016.

[6] A. Al-Hashedi et al., « Ensemble Classifiers for Arabic Sentiment Analysis of Social Network (Twitter Data) towards COVID-19-Related Conspiracy Theories », Appl. Com-put. Intell. Soft Comput., vol. 2022, p. 1-10, janv. 2022, doi: 10.1155/2022/6614730.

[7] M. Al Omari, M. Al-Hajj, A. Sabra, et N. Hammami, « Hybrid CNNs-LSTM Deep Ana-lyzer for Arabic Opinion Mining », in 2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS), Granada, Spain, oct. 2019, p. 364-368. doi: 10.1109/SNAMS.2019.8931819.

[8] L. Yang, Y. Li, J. Wang, et R. S. Sherratt, « Sentiment Analysis for E-Commerce Prod-uct Reviews in Chinese Based on Sentiment Lexicon and Deep Learning », IEEE Access, vol. 8, p. 23522-23530, 2020, doi: 10.1109/ACCESS.2020.2969854.

[9] M. Heikal, M. Torki, et N. El-Makky, « Sentiment Analysis of Arabic Tweets using Deep Learning », Procedia Comput. Sci., vol. 142, p. 114-122, 2018, doi: 10.1016/j.procs.2018.10.466.

[10] H. Saleh, S. Mostafa, A. Alharbi, S. El-Sappagh, et T. Alkhalifah, « Heterogeneous Ensemble Deep Learning Model for Enhanced Arabic Sentiment Analysis », Sensors, vol. 22, no 10, p. 3707, mai 2022, doi: 10.3390/s22103707.

[11] H. A. Galal Elsayed, S. Chaffar, S. Brahim Belhaouari, et H. Raissouli, « A two-level deep learning approach for emotion recognition in Arabic news headlines », Int. J. Comput. Appl., vol. 44, no 7, p. 604-613, juill. 2022, doi: 10.1080/1206212X.2020.1851501.

[12] N. Hicham, S. Karim, et N. Habbat, « An efficient approach for improving customer Sentiment Analysis in the Arabic language using an Ensemble machine learning tech-nique », in 2022 5th International Conference on Advanced Communication Technolo-gies and Networking (CommNet), 2022, p. 1-6. doi: 10.1109/CommNet56067.2022.9993924.

[13] S. Ardabili, A. Mosavi, et A. R. Várkonyi-Kóczy, « Advances in Machine Learning Modeling Reviewing Hybrid and Ensemble Methods », MATHEMATICS & COMPUTER SCIENCE, preprint, août 2019. doi: 10.20944/preprints201908.0203.v1.

[14] O. Sagi et L. Rokach, « Ensemble learning: A survey », WIREs Data Min. Knowl. Dis-cov., vol. 8, no 4, p. e1249, 2018, doi: https://doi.org/10.1002/widm.1249.

[15] Y. Freund et R. E. Schapire, « A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting », J. Comput. Syst. Sci., vol. 55, no 1, p. 119-139, août 1997, doi: 10.1006/jcss.1997.1504.

[16] K. Sarkar, « A Stacked Ensemble Approach to Bengali Sentiment Analysis », in Intelli-gent Human Computer Interaction, Cham, 2020, p. 102-111.

[17] M. A. Omari, « OCLAR: logistic regression optimisation for Arabic customers' reviews », Int. J. Bus. Intell. Data Min., vol. 20, no 3, p. 251-273, 2022, doi: 10.1504/IJBIDM.2022.122177.

[18] M. Hadwan, M. A. Al-Hagery, M. Al-Sarem, et F. Saeed, « Arabic Sentiment Analysis of Users' Opinions of Governmental Mobile Applications », Comput. Mater. Contin., vol. 72, no 3, p. 4675-4689, 2022, doi: 10.32604/cmc.2022.027311.

[19] I. Abu Farha et W. Magdy, « Mazajak: An Online Arabic Sentiment Analyser », in Pro-ceedings of the Fourth Arabic Natural Language Processing Workshop, Florence, Italy, 2019, p. 192-198. doi: 10.18653/v1/W19-4621.

[20] A. Dahou, S. Xiong, J. Zhou, M. H. Haddoud, et P. Duan, « Word Embeddings and Convolutional Neural Network for Arabic Sentiment Classification », p. 11.

[21] M. Kang, J. Ahn, et K. Lee, « Opinion mining using ensemble text hidden Markov mod-els for text classification », Expert Syst. Appl., vol. 94, p. 218-227, mars 2018, doi: 10.1016/j.eswa.2017.07.019.

[22] Ghosh, H., Tusher, M.A., Rahat, I.S., Khasim, S., Mohanty, S.N. (2023). Water Quality Assessment Through Predictive Machine Learning. In: Intelligent Computing and Networking. IC-ICN 2023. Lecture Notes in Networks and Systems, vol 699. Springer, Singapore. https://doi.org/10.1007/978-981-99-3177-4_6

[23] Alenezi, F.; Armghan, A.; Mohanty, S.N.; Jhaveri, R.H.; Tiwari, P. Block-Greedy and CNN Based Underwater Image Dehazing for Novel Depth Estimation and Optimal Ambient Light. Water 2021, 13, 3470. https://doi.org/10.3390/w13233470

[24] G. P. Rout and S. N. Mohanty, "A Hybrid Approach for Network Intrusion Detection," 2015 Fifth International Conference on Communication Systems and Network Technologies, Gwalior, India, 2015, pp. 614-617, doi: 10.1109/CSNT.2015.76.