Engineering Science and Military Technologies
Volume (5) - Issue (1) - Mar 2021

ISSN: 2357- 0954
DOI: 10.21608/JMTC.2021.66626.1172

Original Article

Integrated bidirectional LSTM–CNN model for customers reviews classification

Hossam Elzayady, Mohamed S. Mohamed, Khaled Badran

Department of Computer Engineering, Military Technical College

Keywords:
Bi-directional long shortterm memory (BI-LSTM), deep learning, convolutional neural network (CNN).

Corresponding Author:
Hossam Elzayady, Department of Computer Engineering, Military Technical College, Cairo, Egypt. **Tel:** 0020126828242, **Email:** hossamelzaiade@gmail.com

Abstract
The tremendous increase of Internet users and various social media platforms provide a massive amount of data. Companies are seeking an automated method to assess their customers' satisfaction with their products. Collecting and analyzing opinions and customers' feedback from social media rely on what so called sentiment classification. Several types of research are carried out to investigate opinions in English. As the Arabic language analysis faces many numerous challenges and problems. In our current research, two powerful hybrid deep learning models (CNN-LSTM) and (CNN- BILSTM) are represented. Bidirectional LSTMs are an expansion of conventional LSTMs that can make substantial improvements in sequence classification tasks and identify the most valuable features, CNN is applied. Various data preparation processes are performed, and two regular deep learning models (CNN, LSTM) are implemented to conduct a series of experiments. Experimental results show that the two proposed models have superior performance compared to baselines deep learning models (CNN, LSTM).Furthermore, the (CNN-BI-LSTM) model exceeds the hybrid (CNN-LSTM) model in terms of achieving highest efficiency.

I. INTRODUCTION

Since the advent of the internet, Social media platforms have started to have remarkable attention as a significant way of communication worldwide. In our present time, the simplest way to stay in contact with each other can happen instantly through social media applications. It can also be used to spread different opinions about what is happening all over the globe. As a result of social media's vast usage, it generates a massive amount of posts, criticisms, and reviews. Establishments and organizations are also keen to learn about their clients' opinions on social media platforms; this is one example of the prominent term called "sentiment analysis"[1]. The majority of sentiment analysis work is focused on English data, leaving only a limited amount of research that handled Arabic data. It is believed that this is due to Arabic sources used in processing sentiments and feedbacks being rare[2]. Arabic is not as simple as English in its formation of lexical combinations making sentiment analysis a more difficult challenge. The base of most Arabic words consists of three main letters, and we can build on them to form many words with different meanings[3,4]. However, this does not affect the massive increase of Arabic posts in various aspects on a daily basis, making it essential to run Arabic analysis[5]. Machine learning in sentiment analysis relies heavily on feature engineering[3,5]. As a result, features play a significant role in classification[6]. As a consequence, most approaches pursue suitable features to produce outstanding performance[6,7]. The majority of machine learning algorithms use vectors of fixed-length features; documents can be referred to as vectors of fixed-length features. Bag-of-words is believed to be one of the best techniques to represent each text, primarily due to its efficacy and clarity, however, without concern to word order[5]. This kind of method can lead to misconception in the categorization of sentiments, primarily given the possibility of referring to different suggestions in same-word phrases[8]. The N-gram is used to represent a phrase in another popular form. This way is deemed more superior than the others[5,9]. The input of the classifier can then be interpreted by a specified representation of the input sentence. Currently, it could be assumed that n-gram models take into account the word ordering in simple phrases, and it already encounters the trouble of data sparsity[7]. Deep learning is preferable rather than machine learning as "Feature Engineering" is unnecessary for deep learning. Deep learning features can be retrieved using a completely automated process and with no human expert involvement. Deep learning allows the analysis of multiple layers by turning complex problems into simpler ones to support feature extraction process capabilities. It also shows the high-level features classification needed in various activities[5]. Deep learning

has made significant progress in diverse aspects of Artificial Intelligence (AI). Speech detection and recognition, image captioning, and natural language tasks are highly based on AI[5,8]. In this research, two significant deep learning models (CNN, Bidirectional LSTM) are utilized on customer feedback in text categorization of Arabic to create a sentiment methodological approach. Taking into account the ambiguity of the Arabic language, multiple data preparation steps are carried out. Specific hyper-parameters are used for high precision, as well as some training phases are clearly stated. Outcomes have shown that combined (CNN, bi-directional LSTM) achieves remarkable precision compared to other models.

The remainder of this paper is split into several parts. Section 2 defines the relevant works, while baselines deep learning models used in our proposed method are discussed in Section 3. Section 4 illustrates our suggested combined models. The experimental outcomes and analysis are shown in Section 5. Finally, in Section 6, the conclusion and roadmap for future work are presented.

II. Related Work

A large number of research have been conducted on sentiment analysis. This is largely caused by the gradual expansion of the data of individuals in Social Networks sharing their thoughts, considerations, perspectives, remarks, and everyday life[10]. Various forms of sentiment classifications, applications, methods, and approaches are covered in depth[8]. In 2012, addressing CNNs in the issue relating to picture classification became well known and achieved better execution over various methodologies[11]. In the sentence classification model, CNNs for NLP showed unique results[12]. RNN is a major fundamental model of deep learning, Mikolov et al. presented in 2010 a method of performing RNN model on speech recognition[13]. They show that the n-gram method is dominated by RNN. Using the prior state to figure its existing expression, RNN has different aspects in linguistic layout, which identifies the particular approach in nearly standard languages. The researchers suggested an effective method for sentiment analysis of Arabic languages based on tweet platform in[3], which focused on lexical normalization of the original tweet language. To assess the polarity of each tweet's sentiment, the Support Vector Machines (SVM) classifier for Bag of Words (BOW) representation had the highest sentiment classification accuracy. In[5], to classify customers' attitudes at a particular time, a deep learning methodology is adopted on various Arabic datasets domains. The findings indicate that deep learning efficiency is distinguished from the classic approach to machine learning. DNN achieved an average value of accuracy 90.22%, precision 90.56%, recall 90.90%, and F-measure of 90.68%, compared to other common algorithms of machine learning Naïve Bayes, Decision Tree, and K-Nearest. The researchers in[14] developed an efficient integrated CNN and LSTM model for the English classification of text. The model has a strong potential to be far more precise than the fundamental models (CNN, LSTM).The authors in[24] offer an integrated (CNN-BiLSTM) model applied to articles in French newspapers. They also used Word2vec/Doc2vec embedding. The suggested model was compared to five deep learning models. According to the findings, combined (CNN-BiLSTM) achieves the highest accuracy of 90.66 %. The authors in[22] demonstrate that a deep learning model can outperform a classic machine learning model. On two Arabic datasets, the combination (CNN-LSTM) achieved the highest accuracy of (85.38 %, 86.88 %).

III. Baseline Models

III.1 Convolution neural networks (CNN)

CNNs are a popular type of deep learning structures that is used mainly in the categorization of images, it has also been used to classify texts recently.Figure1 describes the layout of CNN network structure which is used in text categorization. Each sentence is represented as a matrix and each row of matrix represent a word[15]. To make sure that all matrix rows are similar in length, the padding method is followed. CNNs Network includes many processes; first convolution process uses several filters to extract the most substantial features. Then these extracted features are passed on second process which is pooling. The prevalent public technique used in pooling is max pooling; this layer aims to record the maximum value from each feature map[15,16]. Then, the pooled features will be combined as vector by using a fully connected layer, which uses softmax function to get probable value of each class.

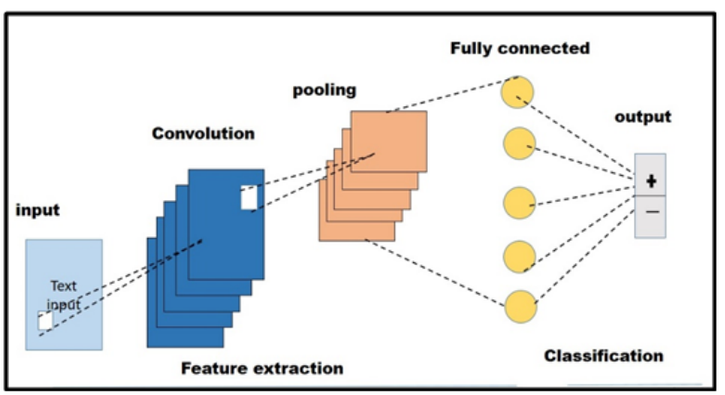

Fig. 1: CNN for Text Classification[16]

III.2 Bi-directional long-short term memory neural networks

Bidirectional LSTM is recognized as an evolved architecture from traditional LSTM that can be used to boost sequence classification problems[15]. The architecture of LSTM is illustrated in Figure 2, there are three crucial parts of LSTMs cell, an input gate, a forget gate, and an output gate. The prime function of input gates is to dominate the new input to the memory. Forget gate is taking charge of preserving values for a particular period of time in memory. Finally, the output gate controls the amount of memory storage required to activate the block[1, 17]. Bidirectional LSTM design permits networks at each time step to gather both forward and backward information about the sequence. The ability to train input via two different methods differs Bidirectional LSTM from conventional LSTM. The first method preserves information from past to future and second preserve information from future to past. There is a potential to preserve information from the future and at any point in time, utilizing the two hidden states combined, preserving information from both past and future is possible. The gates are computed as:

$$G_i^t = \sigma\left(w_i x^t + U_i h^{t-1} + b_i\right) \quad (1)$$

$$G_f^t = \sigma\left(w_f x^t + U_f h^{t-1} + b_f\right) \quad (2)$$

$$G_o^t = \sigma\left(w_o x^t + U_o h^{t-1} + b_o\right) \quad (3)$$

$$C^t = G_f^t \times C^{t-1} + G_i^t \times \tanh\left(W_C x^t + U_C h^{t-1} + b_C\right) \quad (4)$$

$$h^t = G_o^t \times \tanh\left(C^t\right) \quad (5)$$

Each gate's weight matrix is represented by U and W, while bias is given by b. σ and tanh are activation functions

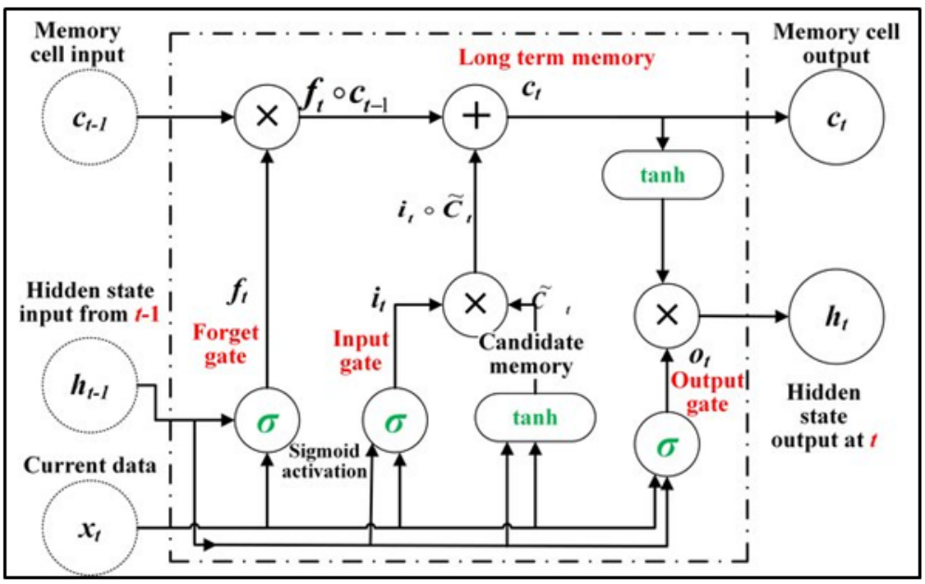

Fig. 2: A repeating LSTM Network[23]

IV. Proposed models

IV.1 CNN-LSTM Model

The design of the suggested (CNN-LSTM) model is displayed in Figure 3, which is comprised of two major parts: (CNN) and (LSTM). The two subsections illustrate how CNN can be used to retrieve higher-level word feature sequences and LSTM to catch long-term correlations across window feature sequences, respectively. The integration starts with a convolution layer that takes word embedding's as inputs. Its output would then be pooled down to more lightweight dimensions and inserted into an LSTM layer. The potential of LSTMs to collect sequential data when considering previous data is one of their strengths. The output vectors from the dropout layer are used as inputs in this layer. Until being moved to a fully connected layer, the LSTM outputs are merged and arranged in a single matrix. The array is converted into a single output in the 0 to 1 range by the fully connected layer.

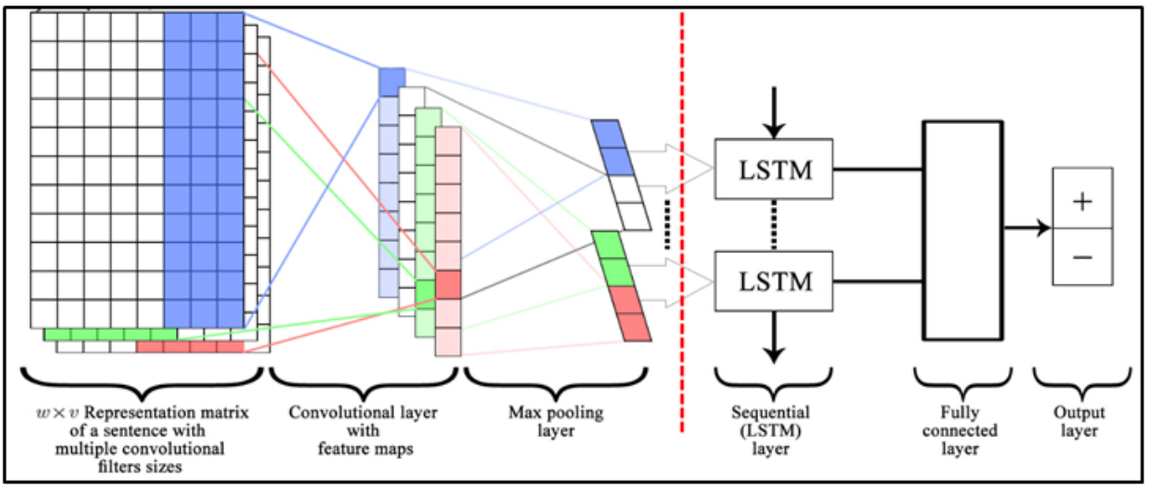

Fig. 3: (CNN-LSTM) proposed model.

IV.2 CNN-BI-LSTM MODEL

Figure 4 demonstrates the CNN-BILSTM model's layout. The proposed model is viewed as an improvement of CNN-LSTM where each LSTM cell is reinforced by two sets of hidden and cell states, one for a forward sequence and the other for a backward sequence. Our proposed model exploits the main features of both LSTM and CNN. In fact, LSTM could accommodate long-term dependencies and overcome the key issues with vanishing gradients. For this reason, LSTM is used when longer sequences are used as inputs. On the other hand, CNN appears able to understand local patterns and position-invariant features of a text. The proposed architecture incorporates different layers. Initially, the embedding layer can convert the input token to vector and pass it through the convolution layer, which applies some filters to upgrade and reduce the data size. Moreover, a layer of max-pooling is added after each filter. Consequently, the outcomes of max-pooling layers are combined to build the input of BILSTM. The results of this stage are the input of a complete fully connected layer, which links each piece of input information with a piece of output information. Eventually, soft max function is used as an activation function to assign classes to each sentence to obtain the required output. Penultimate layer dropout is used to eliminate co-dependencies and reduce overfitting and regularization just as we did in training, by setting activation to 0 for a random proportion p of the hidden units.

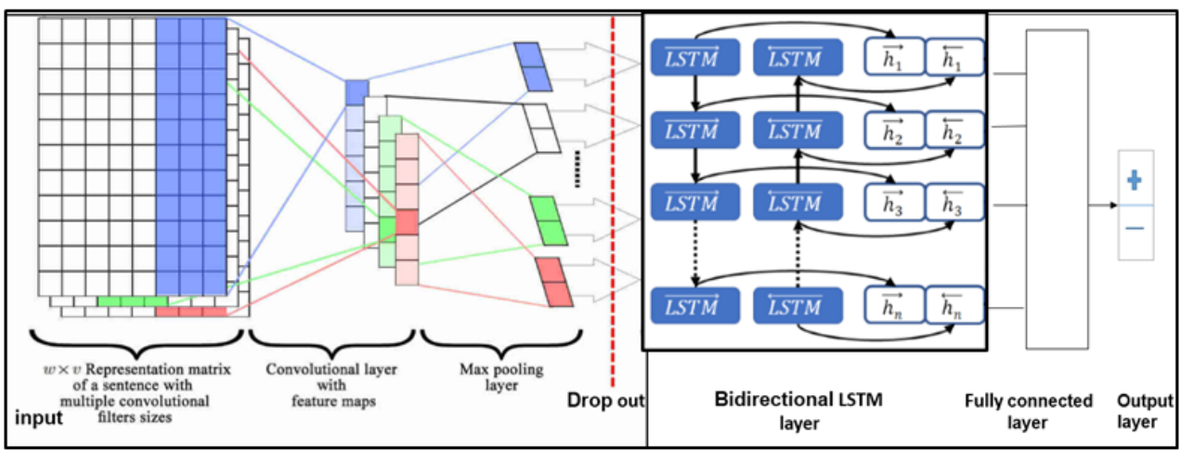

Fig. 4: (CNN-BILSTM) proposed model.

V. Experiments

V.1. Dataset

The LABR dataset has 16449 categorized reviews; in each row, 1 stands for positive review while 0 stands for negative review. The distribution equality between positive class and negative class is taken into account. The dataset includes 8224 positive instances and 8225 negative instances. The book reviews were accumulated by[18, 19] from a general public driven[20]. Figure 5 shows a screenshot from a dataset.

من الكتب الموثره في فكر الشباب المضطرب دينيا .اري انه مرشد مهم ومن ابداعات دكتور مصطفي	1
مش عارفة ليه معجبتنيش المجموعة ده مش عرفة اوصل للي عايزة تقوله الكاتبة بس عموما عجبتني قصة ميلودى رغم انها كئيبة اوى	0
من الكتب الرائعه اللي قريتها في حياتي	1
لم تصل للمستوى الذي كنت أتوقع ! هنالك حلقة مفقودة, مع ذلك.؟.هنالك حقائق رائعة وكلمات جميلة أثارت إعجابي.	0
1من أروع ما قرأت طيلة حياتي كنت في كل صفحة اتحمس للصفحات التالية عزازيل رواية ادبية كتبت بلغة رهيبة بليغة متقنة	1
".لم انتهي منه ,قرات نصفه ,كتاب ممل "	0

Fig. 5: Screen snapshot of sample dataset.

V.2. Rmsprop optimizer

Rmsprop stands for Root Mean Square Propagation; the key concept of this particular optimizer is utilizing different learning rates for each weight were specifying the exact learning rate across all weights is proven to be ineffective. RMSProp divides a gradient by a running average of its recent magnitude[21, 22]. First, the following equation determines the summation:

$$s_t = \gamma s_{t-1} + (1-\gamma)\, g_i^2 \qquad (6)$$

Where s_t refers to the summation of w_i and g_i indicates the gradient of w_i and γ is moving average term while the value of γ is 0.9. Therefore, when gradient is large, it will be reduced. The following equation is used to measure the modified rule:

$$w_{t+1} = w_t - \frac{\alpha}{\sqrt{s_t+\varepsilon}} g_i \qquad (7)$$

Where α is the learning rate and ε is the fuzz factor.

The general architecture of our suggested (CNN-BILSTM) solution, taking into consideration the previously described aspects, is depicted in Figure 6. The Keras library is extensively used in the development of our proposed models. Keras is one of the ultimate APIs for high-level neural networks. It is implemented in Python, and it is essential for plenty of back-end neural network computation engines. The results are determined in accordance with the accuracy, precision, recall, and F1-score values.

$$\text{Accuracy} = \frac{\textbf{CorrectPredictions}}{\textbf{TotalPredictions}} \qquad (8)$$

$$\text{Precision} = \frac{\textbf{Truepositive}}{\textbf{Truepositive}+\textbf{falsepositive}} \qquad (9)$$

$$\text{Recall} = \frac{\textbf{Truepositive}}{\textbf{Truepositive}+\textbf{falsenegative}} \qquad (10)$$

$$\text{F1-score} = 2 \times \frac{(\textbf{Precision} \times \textbf{Recall})}{\textbf{Precision}+\textbf{Recall}} \qquad (11)$$

Before actually moving data to word embedding phase, various data preparation techniques are enforced such as (tokenization, ignore punctuation marks, disregard stop-word, lowercase conversion). For precise accuracy, 5-fold cross-validation is carried out, which is continued until each of the 5 folds is selected as the testing set. Calculating the overall average accuracy of the five testing sets gives the model's accuracy. To implement these tests, data set is segregated, 70 % is chosen for training and 30% for testing. For repeated k-fold cross-validation, sick it learn library is utilized.

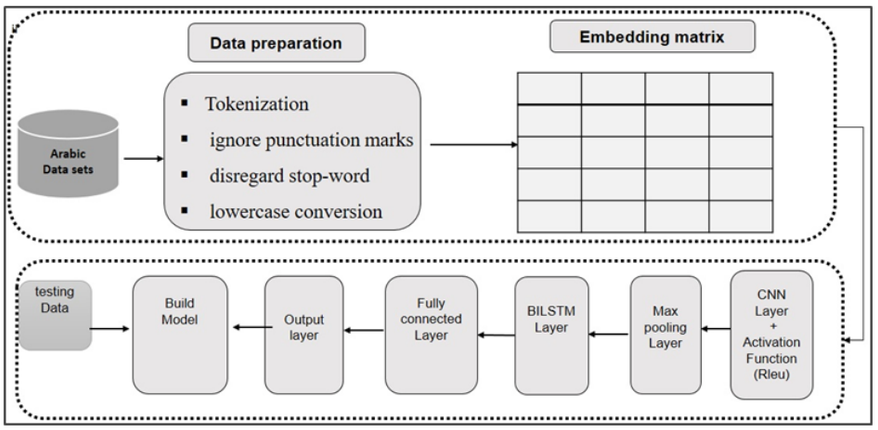

Fig. 6: CNN-BiLSTM general architecture.

In Table 1, all parameters which are used in the training stage are shown. Model validation accuracy is observed for each epoch. The binary cross entropy loss function between both the softmax layer's outputs and their matching labels is reduced.

Table 1: Parameters are chosen for the training phase

embeding dimensions	32
epochs	10
The size of batch	64
filter	64
convolution function	relu
kernel	3
pool size	2
dropout ratio	0.5
loss function	binary_crossentropy
optimizer used	rmsprop
LSTM state dimension	200
word embeddings	not pre-trained

The average dataset validation accuracy of 10 epochs for our models is shown in Figure 7.CNN-BILSTM is performing extremely well. Our combined model achieves 92.4 percent validity precision, beating CNN, LSTM and hybrid (CNN-LSTM) models. However, the average accuracy of each model after 10 epochs is presented in Table.1, following 5 tests that have been conducted.

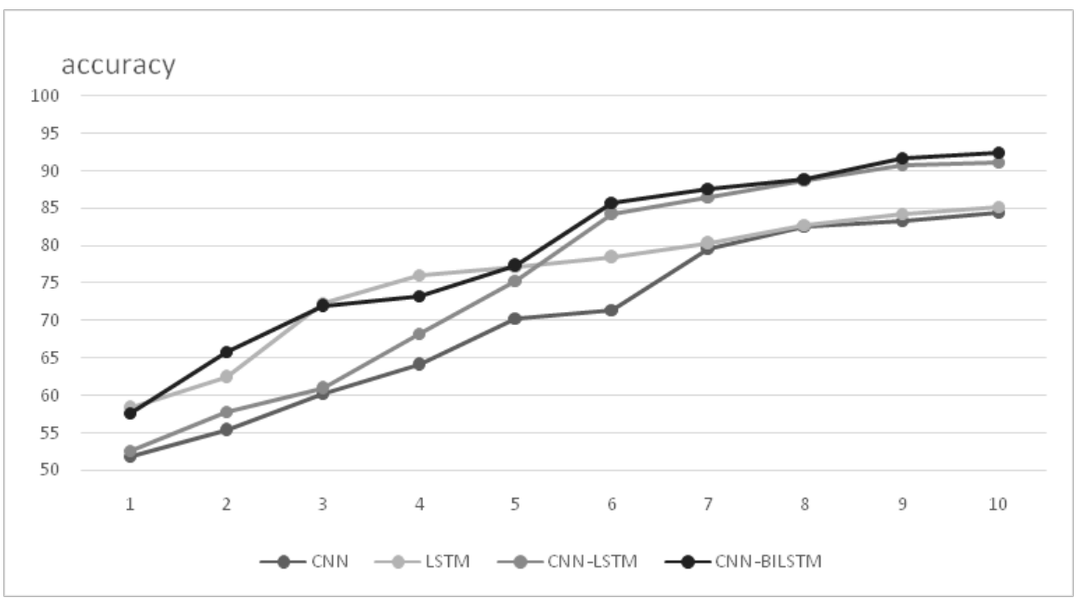

Fig.7: LABR datasetValidation accuracy

The effectiveness of our suggested research methodology to gain the benefit of both CNN in identifying local patterns and the capability of BI-LSTMS to leverage long-term dependencies is shown in Table.2.The results demonstrate CNN-BILSTM model reaches an accuracy of 87.8 excel than both CNN, LSTM and (CNN-LSTM) which reach accuracy of 84.6%, 85.3%, 86.6 %, respectively.Our proposed model (CNN-BILSTM) is evaluated against another model (CNN-LSTM) in [21]. It is noticeable that by adding BILSTM instead of standard LSTM recurrent layer, performance achieves higher accuracy with 1%.

Table 2: Various used models average accuracy, precision, Recall, F1-score

Model	Avg. Accuracy	Avg. Precision	Avg. Recall	Avg. F1-Score
CNN	84.6%	82.9%	82.5%	82.7%
LSTM	85.3%	83.7%	83.2%	83.4%
(CNN-LSTM)	86.6%	84.4%	84.1%	84.2%
(CNN-BILSTM)	87.8%	85.9%	86.2%	86.0%

VI. CONCLUSION

This study proposes using integrated both (Bidirectional LSTM and CNN) models to examine sentiment in Arabic text reviews. In the beginning, reviews are presented by a word vector, then CNN is utilized to get the most relevant features. The BILSTM main objective to gain the context information of the text. Finally, for model enhancement, parameters are tuned. The experimental outcomes confirm the achievability and effectiveness of our proposed model. The proposed model can be improved for future objectives for Arabic categorization, using attention mechanism. We also assume that precision can be increased by using word embedding's such as ELMO and Fast text embedding.

VII. ACKNOWLEDGMENTS

The authors would like to thank Prof. Dr. Mohamed Elshafey and Dr. Ashraf Abosekeen, From Electrical engineering branch, Military technical college for their tremendous help and support.

VII. REFERENCES

[1] Vaateekul, P., and Komsubha, T. (2016, July). A study of sentimen anlysis using deep learning technique on Thai Twitter data. In Computer Scienc and Softwares Engineering (JCSSE), 2016 13th Internationl Joint Conference on (pp. 1-6). IEEE.
[2] Hamad, M., and Al-wadiy, M. (2016). Sentiment analysis for arabic review in social network using machine learning. In Informations Technology: New Generations(pp. 131-139). Springer, Cham.
[3] Alwakd, G., Osmaan, T., and Hughees-Roberrts, T. (2017). Challenge in Sentiment Analysis for Arabic Social Network. Procedia Computers Science, 117, 89-100.
[4] Altowayaan, A. A., andTaao, L. (2016, December). Word embedding for Arabic sentiments analysis. In Big Data (Big-Data), 2016 IEEE International Conference on (pp. 3820-3825). IEEE.
[5] Abdelhadee, N., Solimann, T. H. A., and Ibrahim, H. M. (2017, September). Detecting Twitters Users' Opinion of Arabic Comment During Various Time Episodes via Deep Neurals Networks. In International Conferences on Advanced Intelligent Systems and Informatic (pp. 232-246). Springer, Cham.
[6] Taanag, D., Weei, F., Yangee, N., Zhou, M., Liu, T., and Qin, B. (2014). Learning sentiments-specific word embeddings for twitter sentiments classifications. In Proceedinges of the 52nd Annual Meeting of the Associations for Computational Linguistic (Volume 1: Long Papers) (Vol. 1, pp. 1555-1565).
[7] Lee, Q., andMikolovee, T. (2014, January). Distributed representation of sentence and documents. In International Conference on Machine Learning (pp. 1188-1196).
[8] Medhaet, W., Hasan, A., and Korashy, H. (2014). Sentiment analysis algorithm and application: A surveys. Ain Shames Engineering Journal, 5(4), 1093-1113.
[9] Alomarai, K. M., ElSherife, H. M., and Shaalean, K. (2017, June). Arabic Tweet Sentiments Analysis Using Machine Learning. In International Conference on Industrial, Engineering and Other Application of Applied Intelligent Systemes (pp. 602-610). Springer, Cham.
[10] Paang, B., and Laee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends® in
Information Retrieval, 2(1–2), 1-135.
[11] Krizhevesky, A., Sutskeveers, I., and Hintoan, G. E. (2012). Imagenet classifications with deep
convolutional neurals networks. In Advances in neural information processing systemes (pp. 1097-1105).
[12] Kalchbrener, N., Grefenstetse, E., and Blunsomns, P. (2014). A convolutional neurale networks for modelling sentences. arXiv preprint arXiv:1404.2188.
[13] Mikolovv, T., Kaarafiát, M., Burgett, L., Černoocký, J., and Khudaenpuar, S. (2010). Recurrent neural network based language models. In Eleventh Annual Conference of the International Speech Communications Association.
[14] Soa, P. M. (2017). Twitter sentiments analysis using combined LSTM-CNN models. Eprint Arxiv,1-9.
[15] Roshanfeker, B., Khadvi, S., and Rahmati, M. (2017, May). Sentiment analysis using deep learnings on Persians text. In Electrical Engineering (ICEE), 2017 Iranian Conferencess on (pp. 1503-1508). IEEE.
[16] Senthill Kumrrr, N. K., and Malarvizhi, N. (2020). Bi-directional LSTMs–CNNs combined method for sentiments analysis in part of speech tagging (PoS). International Journal of Speech Technology, 23, 373-380.
[17] Lii, D., and Qiann, J. (2016, October). Text sentiment analysis based on long short-term memory. In 2016 First IEEE International Conference on Computers Communications and the Internet (ICCCI) (pp. 471-475). IEEE.
[18] Altowayaan, A. A., and Taoo, L. (2016, Decembr). Word embeddings for Arabic sentiments analysis. In Big Data (Big Data), 2016 IEEE International Conference on (pp. 3820-3825). IEEE.
[19] Ali, M., and Atiiya, A. (2013, August). Labr: A large scale arabic book review datasets. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistic (Volume 2: Short Papers) (pp. 494-498).
[20] http://www.goodreads.com
[21] Bakthaa, K., and Tripathhy, B. K. (2017, April). Investigations of recurrent neural network in the field of sentiments analysis. In Communication and Signal Processing (ICCSP), 2017 International Conference on (pp. 2047-2050). IEEE.
[22] Zayady, H., Badran, K. M., and Salama, G. I. (2020). Arabic Opinions Mining Using Combined CNN-LSTM Model. International Journal of Intelligent System and Application, 12(4).
[23] Zebin, T., Sperrin, M., Peek, N., and Casson, A. J. (2018, July). Human activity recognition from inertial sensor time-series using batch normalized deep LSTM recurrent networks. In 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 1-4). IEEE.
[24] Rhanoui, M., Mikram, M., Yousfi, S., and Barzali, S. (2019). A CNN-BiLSTM model for document-level sentiment analysis. Machine Learning and Knowledge Extraction, 1(3), 832-847.