Hindawi
Complexity
Volume 2022, Article ID 5217491, 9 pages
https://doi.org/10.1155/2022/5217491

WILEY | Hindawi

Research Article

A Deep Neural Network-Based Approach for Sentiment Analysis of Movie Reviews

Kifayat Ullah,[1] **Anwar Rashad**,[1] **Muzammil Khan**,[1] **Yazeed Ghadi**,[2] **Hanan Aljuaid**,[3] **and Zubair Nawaz**[4]

[1]*Department of Computer and Software Technology, University of Swat, Swat, Pakistan*
[2]*Department of Software Engineering/Computer Science, Al Ain University, Al Ain, UAE*
[3]*Computer Sciences Department, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University (PNU), P.O. Box 84428, Riyadh 11671, Saudi Arabia*
[4]*Department of Data Science, University of the Punjab, Lahore, Pakistan*

Correspondence should be addressed to Kifayat Ullah; kifayat@uswat.edu.pk

Received 25 April 2022; Accepted 31 May 2022; Published 29 June 2022

Academic Editor: Shahzad Sarfraz

The number of comments/reviews for movies is enormous and cannot be processed manually. Therefore, machine learning techniques are used to efficiently process the user's opinion. This research work proposes a deep neural network with seven layers for movie reviews' sentiment analysis. The model consists of an input layer called the embedding layer, which represents the dataset as a sequence of numbers called vectors, and two consecutive layers of 1D-CNN (one-dimensional convolutional neural network) for extracting features. A global max-pooling layer is used to reduce dimensions. A dense layer for classification and a dropout layer are also used to reduce overfitting and improve generalization error in the neural network. A fully connected layer is the last layer to predict between two classes. Two movie review datasets are used and widely accepted by the research community. The first dataset contains 25,000 samples, half positive and half negative, whereas the second dataset contains 50,000 specimens of movie reviews. Our neural network model performs sentiment classification among positive and negative movie reviews called binary classification. The model achieves 92% accuracy on both datasets, which is more efficient than traditional machine learning models.

1. Introduction

The most pleasant passage of time in the modern world is watching movies. People like to give their opinion on movies [1]. Movie reviews are comments or views of persons who have watched the movie. The collection of all these reviews assists the users in knowing whether the movie is worth watching [2].

Sentiment analysis is a field of study in which we examine people's sentiments, opinions, attitudes, and emotions based on some entities like services, products, organizations, topics, events, and their attributes [1]. The objective of sentiment analysis is to find whether these pieces of text convey negative, positive, or neutral opinions.

There are some dissimilarities between sentiment analysis and opinion mining. The first significant difference is that opinion must always have an owner (holder/owner of the comment) and target (about which holder is commenting), whereas sentiment does not have to. Spontaneously, opinion mining is used to find out people's viewpoints (e.g., disagree, agree) on someone or something [3]. Sentiment analysis is used to mine someone's feelings or attitude. If we respond, "I share your sentiment or feeling," this is called sentiment, while if someone's expression can be answered by "I disagree/agree," it is an opinion. It is not required to differentiate between sentiment and opinion in most cases explicitly. This paper considers the words sentiment analysis and opinion mining interchangeably.

To solve the problem of sentiment analysis, three methods are used: lexicon-based, machine learning, and hybrid-based methods [4].

Lexicon-based techniques use either dictionary-based techniques or corpus-based techniques. Dictionary-based methods use dictionary terms, like SentiWordNet and WordNet, and corpus-based approaches use statistical analysis [4].

Machine learning is automatically used to learn from labeled or unlabeled data automatically, called supervised learning or unsupervised learning. At the same time, a hybrid is a combination of both. Machine learning is further divided into traditional machine learning and deep learning.

Machine learning techniques are also known as traditional machine learning techniques, where some of the most popular techniques are SVM (Support Vector Machine), Decision Tree, NB (Naïve Bayes), and RF (Random Forest).

Deep learning mimics the working of the human brain to process data and creates patterns. These patterns are used for the decision-making process. The deep learning approach can learn automatically and get improved from experience without any explicit programming.

This research work proposes a seven-layer deep neural network model for larger datasets. First, the data is converted into word vectors. According to [5], Word2Vec is one of the most powerful techniques that Keras offers in an embedding layer. We use two layers of the convolutional layer. The first 1D convolutional layer is used to extract features from the input data to produce a feature map. The second convolution layer summarizes the features selected by the first convolutional layer. The global max-pooling layer reduces the resolution features of the output and prevents overfitting of the data (Dang, Moreno-García, & De la Prieta, 2020).

The dropout layer is used to solve the problem of generalization and overfitting. This layer randomly drops units from the network, while these units are dropped temporarily, along with all the outgoing and incoming connections [6].

The dense layer uses the loss function on the trained dataset to classify the input features into positive or negative.

The remainder of the paper is organized as follows: Section 1.1 introduces the related research work about movie reviews and machine learning and sentiment analysis, Section 2 describes the proposed seven-layer deep neural network model, Section 3 describes the experiments and results of our proposed model, Section 4 presents discussions, and Section 5 is the conclusion of the research work.

1.1. Related Works. Reference [7] proposed a deep neural network model for sentiment analysis applied to comments of YouTube videos, written in Brazilian Portuguese. The model has six layers. This model achieves an accuracy of 60% to 84% [7].

Reference [6] showed that overfitting is a severe problem in deep neural networks. A deep neural network with large parameters is a powerful but slow machine. The problem of overfitting is handled using the dropout technique. This paper shows that dropout improves the neural network's performance in supervised learning.

There are certain challenges in processing natural languages. In recent years, it has been observed that a neural network is a favorable solution to the challenges of natural language processing [8]. Moreover, according to [9, 10], the deep learning approach with two hidden layers is more accurate than the approach with a single hidden layer.

Reference [11] categorized the sentiment analysis into three levels. The first level is the document level, which is used to categorize the whole document as negative or positive. The second is sentence level, which classifies the sentence. The third is the aspect or feature level that classifies the document or sentence based on some aspect.

Reference [12] proposed an unsupervised-based method in which WordNet dictionary is used to determine opinion words and their antonyms and synonyms. This method is applied to movie reviews classifying each document as negative, positive, or neutral. This model uses POS tagger on the collected reviews, which tags all words of the document. Reference [12] prepared a seed list that holds some opinion words and their polarity. WordNet is used to find its synonyms if the extracted word is not in the seed list. The results have concluded that the proposed model achieves 63% accuracy on document classification using movie reviews.

In [13], the proposed SLCABG model combines the advantages of deep learning and sentiment lexicon. To enhance the sentiment features within the review, a sentiment lexicon is used at the first stage of the model, and then a convolutional neural network and GRU (Gated Recurrent Unit) are used. These layers extract sentiment features and then use the attention mechanism to weigh.

References [14, 15] presented the key challenges faced in sentiment analysis. Reference [16] presented the two most essential comparisons of sentiment analysis. First, it discussed the comparisons between sentiment review structure and analysis challenges. This challenge reveals another factor that challenges faced in sentiment analysis are domain-dependent. Reference [16] shows that the negation challenge is very popular in all types of reviews when there is a minor difference in explicit or implicit meaning. It is also concluded that the nature and review structure present a suitable challenge. Second, it examines the significance of sentiment analysis challenges to improve accuracy. Another hot area of research is the theoretical type of sentiment analysis. These two comparisons are collected from 47 research studies [16].

The WHO (World Health Organization) declared an illness on 11th March 2020 because of COVID-19 (Coronavirus Disease of 2019). As a result, a significant amount of pressure has mounted on each country to assess the cases of COVID-19 and efficiently utilize the available resources, as there was fear, panic, and anxiety as the number of cases increased [17]. More than 24 million people had positive tests worldwide as of 27th August 2020 [17].

Reference [17] extracted facts from tweets related to WHO and COVID-19 and believes that World Health Organization is ineffectual in guiding the public. The authors therein described that two types of tweets were analyzed. First, they gathered tweets from 01-01-2019 to 23-03-2020,

which were around 23,000. These tweets were analyzed and concluded that most of the tweets conveyed negative or neutral sentiments.

The second dataset collected from December 2019 to May 2020, which contained about 226,668 tweets, was analyzed and it was concluded that most of the tweets conveyed positive or neutral sentiments [17]. The authors claimed that the people had posted mostly positive tweets. This claim was then validated using a deep neural network classifier having 81% accuracy.

Reference [18] focused on the multilingual text data of social media to discover the concentration of extremism in sentiment. The authors therein used four classifications, neutral, moderate, low extreme, and high extreme. Reference [18] extracted Urdu data from a different source, which was then authenticated by Urdu domain professionals, and it achieved 88% accuracy. Naïve Bayes and SVM were applied for classification purposes and achieved 82% accuracy.

According to [19], data encoding has a vital role in convolutional neural network training. One of the most popular and simplest encoding methods is one-hot encoding. But when the dataset becomes large, data does not spread the full-label set. Therefore, in one-hot encoding, the relationship between words is independent. However, when the larger dataset is used, the dimension of the word vector will be very large.

Reference [19] had two contributions. First, the authors therein showed that random projections are an effective tool for calculating embedding having lower dimensions. Second, they proposed a normalized eigenrepresentation of the class, which encodes targets with minimal information loss and improves the accuracy of random projections with the same convergence rates.

Reference [20] used an artificial neural network trained on a movie review database with two large lists of negative and positive words. This model achieved 91% accuracy on training and 86.67% accuracy on validation.

The study performed by [6] showed that overfitting is a severe problem in deep neural networks. The deep neural network having an enormous parameter is a powerful but slow machine. This research work addresses the issue of overfitting, which means too much learning will be performed poorly on new data but, on the other hand, there is good performance on the training dataset. These problems are handled using the dropout technique. This paper demonstrates that the neural network performs better by using a dropout on supervised learning. The dropout value may be between 20% and 50%; small value has minimal effect and too large value results in underlearning by the network.

CNN is the most famous technique in computer vision; however, recent studies show that convolutional neural networks perform well on natural language processing. Reference [21] presented Facebook fast-text word embedding to represent words for sentiment analysis instead of word embeddings and trained a convolutional neural network.

Reference [22] proposed a deep neural network model and conducted experiments using a different number of CNN layers and different numbers and sizes of filters. Reference [22] also performed sentiment analysis of Hindi movie reviews, 50% of the dataset was used for training, and the other 50% of the dataset was used for testing the model. The model proposed by [22] achieved 95% accuracy and performed superior to traditional machine learning techniques.

The model proposed by [22] comprises four layers, embedding layer, convolutional layer, global max-pooling layer, and dense layer.

The first layer, that is, the embedding layer, represents a sequence of words as vectors, where the dimension must be less than vocabulary size. Reference [22] used Word2Vec to capture semantic properties. The trained model maps a word to its corresponding vector representation. Softmax probability is used to calculate high-dimensional vectors for every word.

Every hidden neuron accepts a one-word vector. This means that there must be a hidden neuron in the model for every word vector.

Reference [22] used the convolutional layer having "m" kernels/filters to a frame of size "h" over every sentence. These "m" kernels operate in parallel generating several features.

The features map produced by the convolutional layer is given to the global max-pooling layer, which samples these features and generates the local optimum. This layer aggregates information/data and reduces representation.

In [22], the experimental results claimed that the CNN which has two convolutional layers and kernel sizes of 4 and 3 performs better and achieves 95.4% accuracy. In contrast, CNN with three convolutional layers achieves 93.44% accuracy. When the quantity of the convolution layers and kernel size increase, the training time also increases.

Reference [8] reviewed the latest studies and showed that deep learning solves the problems of sentiment analysis and natural language processing. RNN, DNN, and CNN are applied using TF-IDF and word embeddings to many datasets. Word embedding performs well against TF-IDF. The convolution neural network outperforms other models, which present a good balance between CPU runtime and accuracy.

Reference [23] showed that deep learning is better for sentiment analysis. That paper explains the sigmoid function and how weights are learned in neural networks, and a particular convolution layer and pooling operations are used.

2. Methods

Alexandre Cunha proposed a six-layer neural network model for sentiment analysis to mine features and classified the comments [7]. However, he recommended that further work is needed to make this model efficient for analyzing large datasets (Figure 1).

In this research work, to check the performance of [7], we implemented the same neural network model in Python and applied it to larger movie review datasets. Unfortunately, the model does not give satisfactory results. So we

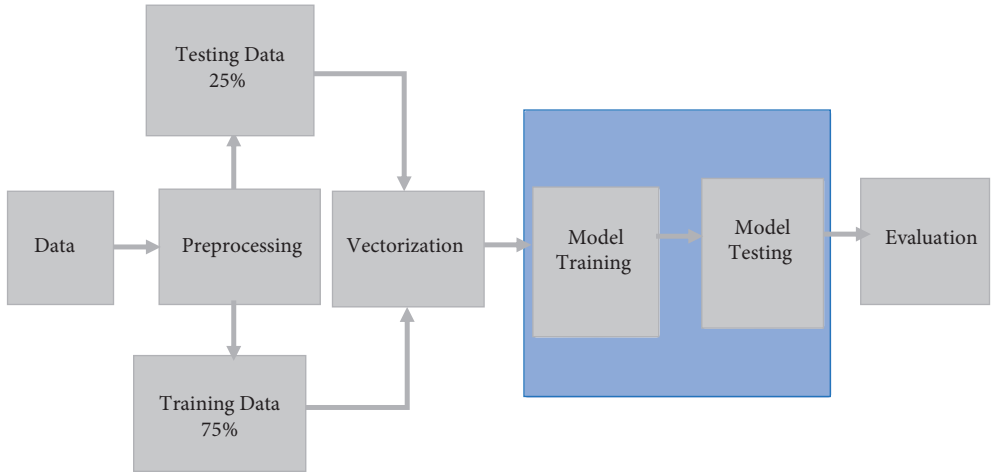

Figure 1: Methodology workflow.

have proposed a seven-layer deep neural network model for larger datasets. Furthermore, in this research work, one more hidden layer (one-dimensional global max-pooling layer) was added with default parameters and some parameters were changed in the proposed model of Alexandre Cunha, which increased the model's accuracy.

Our proposed model has seven layers. The first layer is the embedded layer, followed by two consecutive convolutional layers, a global max-pooling layer, a fully connected layer, a dropout layer, and a dense layer.

2.1. Embedding Layer. The embedding layer is the first layer of the model and is provided with labeled data. The embedding layer requires that the dataset must be cleaned. The datasets are prepared so that one-hot encoding is generated for each word. Size of the vector space must be specified as it is part of the model. We used 256-dimensions.

The word embedding vectors are initialized in the first step with small random vectors. One option is one-hot encoding to map words into word vectors, but there is no relationship between words in one-hot encoding, as each word is independent [19]. Furthermore, the dimensions of the word vector will be very large if the number of words is large. Therefore, the researchers proposed encoding the words into vectors to solve the problem of one-hot encoding [24].

In this study, word embedding is used to convert words into vectors. The vocabulary size is restricted to the top 78,000 most common words for dataset-I and 1,08,000 for dataset-II. 256 dimensions are used and cut off the review after 1200 words, which we select from the experimental results of Table 1. The weights of embedding layers are taken randomly from the dictionary created from the datasets, and then these vectors are adjusted through backpropagation [25].

2.2. Convolution Layer. In the convolution layer, the name "convolution" comes from or means to extract features from the input data, also called filters. In a one-dimensional convolutional layer (1D-CNN), features are extracted from the input data to produce a feature map using a filter or kernel.

The convolutional layer is a feed-forward deep neural network primarily applicable in computer vision, natural language processing, and recommending systems [8]. The objective of a convolutional layer is to extract the most significant features from the input [26].

We used a convolutional layer to effectively extract important features using fewer neurons compared to the dense layer [7]. However, more layers of the convolutional layer will extract more features from the input vectors. Therefore, we have used a two convolutional layers. These convolutional layers have 128 and 64 filters size, respectively.

We have used a 1D convolutional layer. Therefore, it summarizes along with one dimension. Its advantage is that it automatically detects important features without any human supervision.

The second convolutional layer summarizes the features selected by the first convolutional layer because filter size is reduced in this layer.

2.3. Global Max-Pooling Layer. The global max-pooling layer reduces the resolution features in the output and prevents data overfitting [8]. Furthermore, according to [23], pooling layer is used to decrease dimensionality by releasing the number of factors and hence shortening the execution time.

We use the global max-pooling 1D layer. In this layer, the dimensionality of features is decreased without losing key information. The features generated by the convolutional layer are summarized in this layer and thee features of the global region are presented into a feature map. The next layer will perform operations on the summarized features. Now, if there are any variations in the position of input features, the model can identify them. Pooling operations are divided into max-pooling and average pooling. In max-pooling operation, maximum elements are selected from the feature map of the region, which is converted by the filter. In contrast, global max-pooling gives a single value by reducing each channel in the feature map.

Table 1: Effect of the input statement using fixed-length dataset I.

Length of sentence	Accuracy (%)	Precision (%)	Recall (%)	*F*1-Score (%)
10	73.19	73.46	73.56	72.83
2375	87.95	88.05	88.65	87.77
1200	88.89	88.22	87.83	87.35

2.4. Dropout Layer. Multiple nonlinear hidden layers in deep neural networks make the model more expressive which can learn the complicated relationship between input and output [6].

Generalization and overfitting are two significant problems in the neural network. The model's ability to perform well on new data is called generalization. In contrast, when you train the model on large data and the model works well on that data but performs poorly on the test dataset, this problem is called overfitting.

Dropout is a technique that effectively addresses these problems and combines many different architectures of neural networks efficiently [6]. Dropout means dropping visible or hidden units from the network, which are temporarily removed from the network and all its outgoing and incoming connections.

Reference [6] proposed that a typical dropout value for hidden units will be in the range of 0.5 to 0.8 because a shallow value has a very slight effect on the model and a very high value results in underlearning by the neural network.

The dropout layer is used with a dropout rate of 0.3, which we obtained from the experimental results in Table 2. When the dropout layer is used in the neural network, its neurons become less sensitive to specific weights, which will result in a neural network that is less likely to overfit the training data and can be better generalized.

Table 2: Effect of dropout value.

Dropout	Accuracy (%)	Precision (%)	Recall (%)	*F*1-Score (%)
0.3	90.23	90.88	89.91	89.78
0.4	89.81	90.33	89.99	89.52
0.6	89.35	89.45	89.96	89.18
0.8	88.89	88.22	87.83	87.35

Model: "sequential"

Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, 1500, 256)	19968000
conv1d (Conv1D)	(None, 1494, 128)	229504
conv1d_1 (Conv1D)	(None, 1490, 64)	41024
global_max_pooling1d (Global	(None, 64)	0
dense (Dense)	(None, 50)	3250
dropout (Dropout)	(None, 50)	0
dense_1 (Dense)	(None, 1)	51

Total params: 20, 241, 829
Trainable params: 20, 241, 829
Non-trainable params: 0

None

Figure 2: Seven-layer neural network model.

2.5. Fully Connected Layer. Dense layer (densely connected layer) means fully connected layer. A fully connected layer performs primary classification. We have used two dense layers, one hidden layer, and another is the last layer. The hidden dense layer will return an array of 50 probability scores. It takes a feature vector as input and gives an output of a 50-dimensional vector. At the same time, the last dense layer is used to classify features of the input into various classes.

The seven-layer sequential neural network model is shown in Figure 2.

3. Experiments

This section evaluates the neural network model for sentiment analysis.

3.1. Data Collections. There are several sources for movie review datasets like GitHub, Kaggle.com, UCI (machine learning repository), and IMDb. We extracted movie review datasets from IMDb. The critics frequently use movie rating websites like IMDb to post remarks and rate movies, which assist users in deciding whether to watch the movie or not. In this study, we have used two datasets on IMDb, an online database of information about movies. These datasets are available and broadly accepted by researchers [8]. The first dataset (dataset I) is available on Kaggle.com [27], which contains IMDb movie reviews from the audience, with 25,000 samples having half positive and half negative [28]. The second dataset (dataset II) is also available on Kaggle.com [29], and it is a binary classification dataset containing 50,000 reviews of movies.

3.2. Preprocessing. To improve the quality of sentiment analysis and reduce errors and inconsistencies in sentiment analysis, we need to clean the data [1]. We observed from the datasets that they contain HTML tags and punctuations. Therefore, we removed the HTML tags and special characters from these datasets. After removing punctuations, which result in single characters that make no sense, we removed all the strings having single characters. We replaced them with a space that creates multiple spaces in our dataset. Then we removed multiple spaces from our text, and, finally, we converted the text to lowercase. After cleaning, each dataset is divided into two pairs, training and testing; 75% of the dataset is used to train the proposed model and the remaining 25% for testing. In dataset I, from 25,000 samples, 18,750 sequences or samples are used for training, while the

remaining 6,250 sequences or samples are used for testing. Meanwhile, in dataset II, 37,500 sequences or samples are used for training and 12,500 sequences or samples are used for testing from 50,000 samples.

3.3. Performance Metrics. The evaluation metric of the model used in this research work is accuracy. The parameters of the accuracy are defined as follows:

TP: The total number of reviews is is categorized as positive, and the reviews are positive.

FP: The total number of reviews is categorized as negative, and the reviews are positive.

TN: The total number of reviews is categorized as negative, and the reviews are negative.

FN: The total number of statements/reviews is categorized as positive, and the reviews are negative.

Accuracy is calculated by taking the ratio between the predicted reviews and the total number of reviews.

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})}. \quad (1)$$

Precision is calculated by taking the ratio between the reviews predicted correctly as positive and the total number of predictable positive reviews.

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})}. \quad (2)$$

Recall is calculated by taking the ratio between the reviews that are predicted correctly as positive reviews to all of the reviews in that class.

$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})}. \quad (3)$$

*F*1-Score is calculated by taking the weighted average between recall and precision.

$$F1 - \text{Score} = \frac{(2 * \text{Precision} * \text{recall})}{(\text{Precision} + \text{recall})}. \quad (4)$$

3.4. Experiment Results. These experiments are performed using datasets I and II. We took the largest review, average review, and smallest review length to conduct experiments. The parameters of the proposed model are shown in Table 3. We record the experimental results in Tables 1, 4, and 5. We discovered from the experimental results that using the longest review and average review length as a fixed length of the input gives better results compared to the smallest review length. Furthermore, the average length review is more effective in terms of processing speed and accuracy, which affects the performance of the proposed neural network model. The experiments in Table 5 are performed on the six-layer model proposed by [7]. It is shown in Table 5 that the accuracy of the six-layer model is less than that of our proposed model.

Table 3: Parameters of the model.

Parameters	Values
The length of the input sentence	1200
The dimension of the word vector	256
The thesaurus size	78000, 108000
The size of the convolution kernel	7×5
The number of hidden neurons in the convolution layer	128
Dropout	0.3

Table 4: Effect of the input statement using fixed-length dataset II.

Sentence length	Accuracy (%)	Precision (%)	Recall (%)	*F*1-Score (%)
6	73.28	74.74	71.40	72.28
2375	90.85	91.03	90.93	90.47
1200	91.04	91.13	91.09	90.66

Table 5: Effect of the input statement using fixed-length six-layer model.

Length of sentence	Accuracy (%)	Precision (%)	*F*1-Score (%)
10	68.41	66.51	89.35
2375	50.28	50.05	99.78
1200	52.39	53.97	99

The generalization performance of the neural network model is improved by using a dropout layer. Different dropout values are used in the experiment. It has been observed from the experiments that when the dropout value is set to 0.3, the model's performance is optimum. The results of the experiments are shown in Table 2. The experimental result of Table 2 is on the seven-layer model and the experimental results of Table 6 are on the six-layer model [7], whose accuracy is not more than 52.39%.

It has also been observed from the experiments that the model's performance is affected by the number of iterations. When we increase the number of iterations, the performance of the model also improves. It is presented in Tables 7 and 8 and also shown in Figures 3 and 4. The experiments in Table 9 are performed on the six-layer model, with accuracy not more than 55%.

The testing accuracy, recall, precision, and *F*1-Score are shown in Tables 10 and 11.

From Tables 10 and 11, it is noted that the model achieved a final accuracy of 91.18% on dataset I and 91.98% accuracy on dataset II. Meanwhile, from Table 12, it is noted that the six-layer model achieved a final accuracy of 53.70%, which is less than that of our proposed model.

4. Discussions

We proposed a seven-layer deep neural network model for sentiment analysis to classify movie reviews in this research work. Before the word vector is input into the model, the data is preprocessed to improve the quality of sentiment analysis. We remove HTML tags, punctuation marks, and spaces in preprocessing as they do not contain any

Table 6: Effect of dropout value.

Dropout	Accuracy (%)	Precision (%)	*F*1-Score (%)
0.3	52.39	53.97	99
0.4	52.10	53.67	100
0.6	52.25	53.18	100
0.8	51.74	52.08	100

Table 7: Effect of the iterations on the model's dataset I.

Epoch	Accuracy (%)	Precision (%)	Recall (%)	*F*1-Score (%)
1	89.99	90.58	90.01	89.52
2	97.13	97.29	97.01	97.02
3	99.63	99.63	99.64	99.62
4	99.89	99.92	99.88	99.89

Table 8: Effect of the iterations on the model's dataset II.

Epoch	Accuracy (%)	Precision (%)	Recall (%)	*F*1-Score (%)
1	91.04	91.13	91.09	90.66
2	96.84	97.00	96.75	96.74
3	99.18	99.16	99.20	99.14
4	99.42	99.44	99.38	99.38

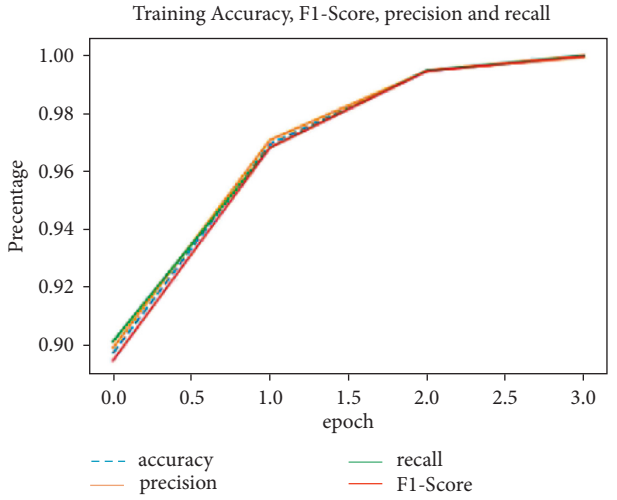

Figure 3: Accuracy on dataset I.

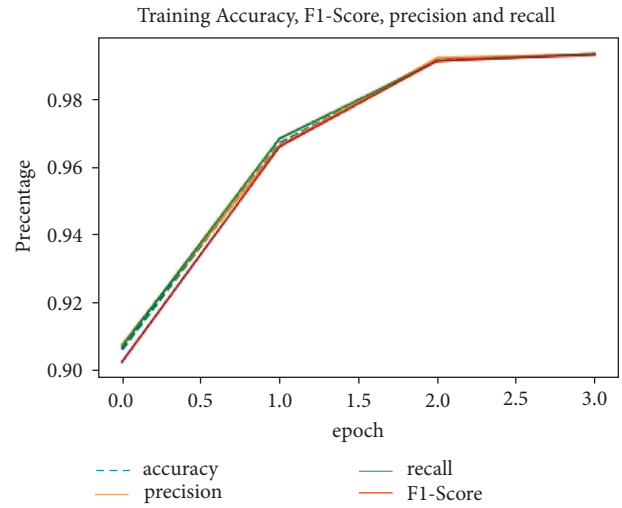

Figure 4: Accuracy on dataset II.

Table 9: Effect of the number of iterations.

Epoch	Precision (%)	Accuracy (%)	*F*1-Score (%)
1	54.09	52.48	100
2	54.59	53.14	100
3	54.58	53.65	100
4	54.93	53.87	100

Table 10: Test results of the model's dataset I.

Epoch	Accuracy (%)	Recall (%)	*F*1-Score (%)	Precision (%)
Testing	92.18	92.53	91.94	91.79

Table 11: Test results of the model using dataset II.

Epoch	Accuracy (%)	Precision (%)	Recall (%)	*F*1-Score (%)
Testing	91.98	94.14	89.93	91.72

Table 12: Test results using dataset I.

Epoch	Accuracy (%)	*F*1-Score (%)	Precision (%)
Testing	53.70	100	52.14

information. For feature selection, we have different options, but the most effective method is word embedding. So we used word embedding (embedding layer), which is used to convert text into vectors so that words with similar meanings get a closer vector.

In addition, we found that the length of the input review affects the model's performance. Therefore, we select small, average, and maximum review lengths, where the average review length gives better results on the model. After the embedding layer, two consecutive convolution layers are used to extract the essential features from the data. We use the convolution layer because it extracts features from fewer neurons compared to the traditional dense layer. The second convolutional layer summarizes the features selected by the first convolutional layer. Then we use the global max-pooling layer to decrease dimensionality without losing key features and prevent the model from overfitting data. A fully connected layer takes the features vector generated by the global max-pooling layer as input and provides an array of probability as output. A dropout layer is used to randomly drop hidden layer neurons to reduce overfitting and improve generalization error. The value of the dropout layer is set to 0.3 because when the value of dropout is set to 0.3, the model's performance is optimum. Finally, a dense layer is used to classify sentiment features as positive or negative.

5. Conclusion

The model proposed by [7] is a deep neural network consisting of six layers applied to two video comments on YouTube. The accuracy of this model is in the range of 60% to 84% [7]. It was suggested that this model would be applied to a larger dataset. We applied the same model to larger

datasets of movie reviews, but it gives 55% accuracy. We added one more layer called the global max-pooling layer in the same model as Alexandre Cunha and changed some parameters, improving the model's accuracy from 55% to 92%. The accuracy of the six-layer model proposed by [7] is less because the dataset used is large, and every review of the dataset is about 1500 words (cut-off review).

We have successfully implemented a deep neural network with seven layers on movie review data. Our model achieves accuracy of 91.18%, recall of 92.53%, *F*1-Score of 91.94%, and precision of 91.79% on dataset I, and, on dataset II, the model achieves accuracy of 91.98%, precision of 94.14%, recall of 89.93%, and *F*1-Score of 91.72%.

Data Availability

The code and data used to support the findings of this study have been deposited in the GitHub repository and are available at https://github.com/asifntu/sentimentanalysis. The readers can easily follow the steps to reproduce the study.

Conflicts of Interest

The authors have no conflicts of interest to report regarding this study.

Acknowledgments

Princess Nourah bint Abdulrahman University Researchers Supporting Project no. PNURSP2022R54, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

References

[1] D. S. Sisodia, S. Bhandari, N. K. Reddy, and A. Pujahari, "A comparative performance study of machine learning algorithms for sentiment analysis of movie viewers using open reviews," in *Performance Management of Integrated Systems and its Applications in Software Engineering*, pp. 107–117, Springer, Salmon Tower Building, NY, USA, 2020.

[2] B. Lakshmi Devi, V. Varaswathi Bai, S. Ramasubbareddy, and K. Govinda, "Sentiment analysis on movie reviews," in *Emerging Research in Data Engineering Systems and Computer Communications*, pp. 321–328, Springer, Salmon Tower Building New York City, 2020.

[3] W. Zhang, M. Xu, and Q. Jiang, "Opinion Mining and Sentiment Analysis in Social media: Challenges and Applications," in *Proceedings of the 5th International Conference, HCIBGO 2018, Held as Part of HCI International 2018*, Las Vegas, NV, USA, July 15-20, 2018.

[4] B. Bhavitha, A. P. Rodrigues, and N. N. Chiplunkar, "Comparative Study of Machine Learning Techniques in Sentimental Analysis," in *Proceedings of the 2017 International Conference on Inventive Communication and Computational Technologies (ICICCT)*, Coimbatore, India, 10-11 March 2017.

[5] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," *Advances in Neural Information Processing Systems*, vol. 26, 2013.

[6] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[7] A. A. L. Cunha, M. C. Costa, and M. A. C. Pacheco, "Sentiment analysis of youtube video comments using deep neural networks," *International Conference on Artificial Intelligence and Soft Computing*, China, 2019.

[8] N. C. Dang, M. N. Moreno-García, and F. De la Prieta, "Sentiment analysis based on deep learning: a comparative study," *Electronics*, vol. 9, no. 3, p. 483, 2020.

[9] A. Alharbi and O. Sohaib, "Technology readiness and cryptocurrency adoption: PLS-SEM and deep learning neural network analysis," *IEEE Access*, vol. 9, pp. 21388–21394, 2021.

[10] O. Sohaib, W. Hussain, M. Asif, M. Ahmad, and M. Mazzara, "A PLS-SEM neural network approach for understanding cryptocurrency adoption," *IEEE Access*, vol. 8, pp. 13138–13150, 2020.

[11] B. J. Liu, *Sentiment analysis and opinion mining*, vol. 5, no. 1, pp. 1–167, 2012.

[12] R. Sharma, S. Nigam, and R. J. Jain, "Opinion Mining of Movie Reviews at Document Level," 2014, https://arxiv.org/abs/1408.3829.

[13] L. Yang, Y. Li, J. Wang, and R. S. Sherratt, "Sentiment analysis for E-commerce product reviews in Chinese based on sentiment lexicon and deep learning," *IEEE Access*, vol. 8, pp. 23522–23530, 2020.

[14] R. Sujata, K. J. Parteek, and R. i. C. Science, "Challenges of Sentiment Analysis and Existing State of Art," 2014, https://www.researchgate.net/publication/308331478_Challenges_of_Sentiment_Analysis_and_Existing_State_of_Art.

[15] G. Vinodhini and R. J. I. J. Chandrasekaran, *Sentiment analysis and opinion mining: A Survey*, vol. 2, no. 6, pp. 282–292, 2012.

[16] D. M. E.-D. M. Hussein, "A survey on sentiment analysis challenges," *Journal of King Saud University - Engineering Sciences*, vol. 30, no. 4, pp. 330–338, 2018.

[17] K. Chakraborty, S. Bhatia, S. Bhattacharyya, J. Platos, R. Bag, and A. E. Hassanien, "Sentiment Analysis of COVID-19 tweets by Deep Learning Classifiers—A study to show how popularity is affecting accuracy in social media," *Applied Soft Computing*, vol. 97, Article ID 106754, 2020.

[18] M. Asif, A. Ishtiaq, H. Ahmad, H. Aljuaid, and J. Shah, "Sentiment analysis of extremism in social media from textual information," *Telematics and Informatics*, vol. 48, Article ID 101345, 2020.

[19] P. Rodríguez, M. A. Bautista, J. Gonzàlez, S. Escalera, and V. Computing, "Beyond one-hot encoding: lower dimensional target embedding," *Image and Vision Computing*, vol. 75, pp. 21–31, 2018.

[20] Z. Shaukat, A. A. Zulfiqar, C. Xiao, M. Azeem, and T. Mahmood, "Sentiment analysis on IMDB using lexicon and neural networks," *SN Applied Sciences*, vol. 2, no. 2, p. 148, 2020.

[21] I. Santos, N. Nedjah, and L. de Macedo Mourelle, "Sentiment analysis using convolutional neural network with fastText embeddings," in *Proceedings of the 2017 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, Arequipa, Peru, 08-10 November 2017.

[22] S. Rani and P. Kumar, "Deep learning based sentiment analysis using convolution neural network," *Arabian Journal for Science and Engineering*, vol. 44, no. 4, pp. 3305–3314, 2019.

[23] K. Chakraborty, S. Bhattacharyya, R. Bag, and A. A. Hassanien, "Sentiment analysis on a set of movie reviews using deep learning techniques," *Social Network Analytics:*

Computational Research Methods and Techniques, vol. 127, 2018.

[24] M. Giatsoglou, M. G. Vozalis, K. Diamantaras, A. Vakali, G. Sarigiannidis, and K. C. Chatzisavvas, "Sentiment analysis leveraging emotions and word embeddings," *Expert Systems with Applications*, vol. 69, pp. 214–224, 2017.

[25] F. Chollet, *Deep Learning with Python*, Simon & Schuster, Manhattan, New York City, 2017.

[26] Y. Chen, J. Wang, X. Chen, A. K. Sangaiah, K. Yang, and Z. Cao, "Image super-resolution algorithm based on dual-channel convolutional neural networks," *Applied Sciences*, vol. 9, no. 11, p. 2316, 2019.

[27] L. N, "Manhattan," 2019, https://www.kaggle.comhttps://www.kaggle.com/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews.

[28] E. Ahmed, M. A. U. Sazzad, M. T. Islam, M. Azad, S. Islam, and M. H. Ali, "Challenges, Comparative Analysis and a Proposed Methodology to Predict Sentiment from Movie Reviews Using Machine Learning," in *Proceedings of the 2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC)*, Chirala, Andhra Pradesh, India, 23-25 March 2017.

[29] L. Andrew, R. E. D. Maas, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts, "Learning Word Vectors for Sentiment Analysis," 2011, https://www.kaggle.com/c/word2vec-nlp-tutorial/data.