Gümüşhane University Journal of Science

GUFBD / *GUJS* (2025) 15(2): 563-576 doi: 10.17714/gumusfenbil.1653072 Research Article

CBLTwitter: Twitter disaster detection analysis using CNN-BiLSTM deep learning methods

CBLTwitter: CNN-BiLSTM derin öğrenme yöntemlerini kullanarak Twitter felaket tespiti

Halit ÇETİNER*[1], Hakan YÜKSEL[1]

[1]*Isparta University of Applied Sciences, Vocational School of Technical Sciences, Department of Computer Technology, 32000, Isparta*

• Received: 07.03.2025 • Accepted: 15.05.2025

Abstract

Twitter, one of the social media platforms, is one of the reliable sources that allows everyone to express their thoughts and ideas online. In this article, we focus on analysing and analysing the text content of tweets on the Twitter platform in extraordinary situations such as possible disasters or disasters. As a result of real-time information from the Twitter platform, it is possible to help people in possible disaster situations and automatically direct emergency teams. In order to prepare the ground for the realization of these possible scenarios, it is necessary to perform high performance classification by identifying disaster-related content from thousands of raw text content. In this paper, we propose a CBLTwitter model that classifies disasters by increasing the weight scores of their significant values that can capture local patterns and contextual dependencies in raw tweet information. The proposed CBLTwitter model investigates the effectiveness of a contextual word embedder called Bidirectional Encoder Representations from Transformers (BERT) in predicting disasters from Twitter data. In addition, BERT results are compared with the results obtained from independent word embedding methods called Word2Vec and Global Vectors for Word Representation (GloVe). As a result, the proposed CBLTwitter model of the BERT word embedder in disaster prediction, which consists of an attention-layer Convolutional Neural Network (CNN) and Bidirectional Long Short Term Memory (BiLSTM) architectures, provided performance results competitive with the literature.

Keywords: Attention mechanism, BERT, BiLSTM, CNN, NLP, Twitter data

Öz

Sosyal medya platformlarından biri olan Twitter, herkesin düşünce ve fikirlerini çevrimiçi olarak dile getirmesini sağlayan güvenilir kaynaklardan biridir. Bu makalede, Twitter platformundaki tweet içeriklerindeki olası afet ya da felaket gibi olağanüstü durumlardaki metin içeriklerinin incelenmesi ve analizine odaklanılmıştır. Twitter platformundan alınan gerçek zamanlı bilgiler neticesinde olası felaket durumlarında insanlara yardımcı olmak ve acil durum ekiplerini otomatik olarak yönlendirme yapmak mümkündür. Belirtilen olası senaryoları gerçekleştirilmesine zemin hazırlayabilmek için binlerce ham metin içeriği içerisinden felaket ile ilgili içerikleri tespit ederek yüksek performans seviyesine sahip sınıflandırma gerçekleştirmek gerekmektedir. Bu makalede, felaketler hakkında karar verebilen ham tweet bilgileri içerisindeki yerel kalıpları ve bağlamsal bağımlılıkları yakalayabilen önemli değerlerinin ağırlık puanlarını artırarak sınıflandırma yapan CBLTwitter modeli önerilmiştir. Proposed CBLTwitter modeli, Twitter verilerinden felaket tahmin etmede Bidirectional Encoder Representations from Transformers (BERT) adlı bağlamsal kelime yerleştiricinin etkinliğini araştırmaktadır. Bunların yanısıra BERT sonuçlarının, Word2Vec ve Global Vectors for Word Representation (GloVe) adlı bağımsız kelime yerleştirme yöntemlerinden elde edilen sonuçlar ile karşılaştırılması yapılmaktadır. Sonuç olarak felaket tahmininde BERT kelime yerleştiricinin attention katmanlı Convolutional Neural Network (CNN) ve Bidirectional Long Short Term Memory (BiLSTM) mimarilerinden oluşan proposed CBLTwitter modeli literatür ile rekabet edebilir performans sonuçları sağlamıştır.

Anahtar kelimeler: *Dikkat mekanizması, BERT, BiLSTM, CNN, NLP, Twitter verisi*

*Halit ÇETİNER; halitcetiner@isparta.edu.tr

1. Introduction

Social media, which has been used extensively in recent times, has enabled everyone to express their thoughts and opinions online. Through the thoughts and opinions of social media users, it is possible to understand what a company, a country or a customer thinks. In addition, social media analysis is used as an effective tool to understand public opinion on political issues (Vadivukarassi et al., 2018; Birjali et al., 2021). In order for text data obtained from social media to be used by deep learning approaches, the texts must be converted into digital vectors (Khan & Yairi, 2018). Word embeddings are an important part of NLP, supporting the process of extracting information from text data and grouping related document content in the process of converting texts into digital vectors (Semary et al., 2023). In many NLP problems, from multi-label text classification to sentiment analysis, the word embedding layer can make a dramatic difference (Çetiner, 2022). Although word embeddings cannot capture contextual information in text analysis, they can change its meaning (Semary et al., 2023). Word embeddings such as ROBERTa and BERT are approaches developed to provide a solution to the mentioned problem. It is important that these approaches automatically perform sentiment analysis on thousands of tweet content obtained from social media in order to provide meaningful insights (Tan et al., 2022). Performing sentiment analysis in terms of long distance dependencies is a difficult problem given the lexical diversity of texts (Tan et al., 2022). There is a need to study whether array models or transformer-based models are better at processing texts with long-distance dependencies. To address this gap, they have constructed models that combine both array models and transformer-based models (Vaswani, 2017; Tam et al., 2021; Tan et al., 2022). RoBERTa models are structures that perform word and sub-word tokenization and word embedding tasks in word embedding. These models are based on the optimized BERT approach from the transformer family.

Many social media platforms such as Facebook, Instagram, blogs, reviews, Twitter, news sites, shopping sites, etc. have to give space to people's social, political or commercial opinions and reviews. The infrastructures of such structures are based on people communicating, interacting or influencing. In the case of Twitter, the number of users of the Twitter platform increased by 190 million from 2012 to 2020 (Tam et al., 2021). On the Twitter platform, there are 145 million users who tweet on a daily basis for all people in the world to see. In order to extract meaningful data from these tweets, it is necessary to discover hidden meaningful values that will determine contextual polarity (Alami & Elbeqqali, 2015; Zhao et al., 2016). In order to determine contextual polarity, word clues extracted from the sentence context need to be analysed and processed in detail (Shaik et al., 2023). From companies in the business world to political party executives, there is a need for people's feedback on many issues such as products and services. On the Twitter platform, it is possible to measure public reaction to political events through policy analysis. Considering the number of tweets, it is not possible to analyse each of them manually and reach a decision. For the reasons mentioned above, there is a need for new methods to automatically analyse ready-made datasets consisting of thousands of tweet content.

Although statistical-based machine learning-based methods have been developed for extracting meaningful data from raw texts and automatic analysis of related texts, it is reported that the desired performance cannot be achieved and these methods fail in complex text classification problems (Neethu & Rajasree, 2013; Huang et al., 2017). There are Long Short Term Memory (LSTM) based deep learning models that can produce meaningful results in complex texts as well as in textual content with high semantic length. Algorithms developed in different architectures based on renewal neural networks, especially LSTM, enable remarkable results in NLP. These architectural models are recommended for their ability to represent textual data in multiple and consecutive layers (Tam et al., 2021). Deep learning based methods provide good performances in classification processes with layered representation of data without the need for more time and resource consuming feature extraction methods (Çetiner, 2024). There are researches that use Convolutional Neural Network (CNN) based architectural models instead of recurrent neural networks for feature extraction and measurement of response from temporal and spatial data (Yeboah & Baz Musah, 2022; Sitaula & Shahi, 2024; Yang & Li, 2024). Integrating BiLSTM architecture and CNN models, which are very successful in capturing the contextual power of texts, can provide effective results in analysing Twitter texts. In this approach, we aim to improve the analysis performance of Twitter texts by using transformer-based BERT and GloVe and Word2Vec word embedding layers of BiLSTM and CNN architectures.

The stated aims and objectives of this article and its main contributions to the literature are presented below.

- In order to numerically represent the tweet texts after certain pre-processing methods, the word embedding techniques BERT, GloVe, Word2Vec are used separately. These models are unsupervised word vectors

that are effective in capturing word semantics since they are built with a large pre-trained word dataset. BERT, GloVe, and Word2Vec word embedding models are applied to verify the effectiveness of the proposed CBLTwitter model.
- With the integration of CNN architectures and BiLSTM architecture from renewal neural networks, it is possible to capture long distance dependencies by obtaining local features.
- The proposed CBLTwitter model was compared with the studies in the literature with the dataset used in this paper. The accuracy of the Proposed CBLTwitter model is 6% better than other studies in the literature.

The following organization of the paper consists of 4 sections. Section 2 presents the theoretical background of previous work that analyses Twitter text content. Section 3 presents the dataset used in the paper and a detailed description of the proposed CBLTwitter model. Section 4 presents a comparison of the proposed CLTwitter model with the performance results obtained from classical CNN and BiLSTM architectures. Section 5 concludes the paper by presenting recommendations for further improvement of the results obtained in this paper.

2. Related works

In text processing, multilayer perceptron structures can perform complex classification operations involving the classification of texts. CNN architectures can be used in computer vision problems such as image processing, voice recognition, speech recognition or retinal recognition, and can be used to find solutions to NLP problems (Kowsher et al., 2021; Çetiner, 2023; Kishwar & Zafar, 2023; Priya & Deepalakshmi, 2023). Feng et al. conducted a study analysing Weibo short texts containing short text information to analyse the changing emotions depending on the specific event and time (Feng et al., 2024). In their work, it is seen that CNN and BiLSTM architectures are used together with attention mechanisms. In the study where the attention mechanism was used to focus on important words in the text, it was determined that the focus was on obtaining semantic and temporal information of the data. Kim and Lee developed an approach combining BiLSTM and CNN structures with transformer structure on two different audio datasets, Emo-DB and Ravdess, to perform sentiment analysis from speech (Kim & Lee, 2023). In their work, CNN captures the spatial details of the sounds, while the BiLSTM transformer structure captures the speech patterns. Wankhade et al. developed a new approach based on CNN and BiLSTM by combining local and global context dependencies with text pre-processing (Wankhade et al., 2024). It is seen that multiple attention mechanism is used to benefit from the synergistic power of CNN and BiLSTM architectures. They tried to capture the emotions in the texts with the attention mechanism. They focused on capturing local patterns with CNN and contextual dependencies with BiLSTM. Sadr and Soleimandarabi have developed a work that tries to overcome the disadvantages of CNN architectures that require a large amount of training data to determine the polarity of a sentence and depend too much on hyperparameters (Sadr & Nazari Soleimandarabi, 2022). In trying to improve the disadvantages of the CNN architecture, word embedding approaches are used to better represent the words as vectors.

When the studies conducted in text analysis in recent years are examined, it is seen that there are studies based on the attention mechanism. These studies try to capture the contextual information of important characters in texts. In this study, inspired by the aforementioned studies, the effects of three different word embedding approaches in a model with correctly realized pre-processing methods on an approach with CNN and BiLSTM architecture are determined. At the same time, as a result of experimental studies, the location of the attention layer that weights the words according to their importance weights was determined. As a result of 18 layers of deep learning layers, an approach that can compete with the studies in the literature is proposed.

3. Material and methods

Word2Vec, GloVe and BERT methods were used separately for word representation to provide a better representation of sentence semantics obtained from Twitter text analysis. The proposed CBLTwitter model was kept the same on the target task and the effect of the differences in word embedding methods was interpreted. While Word2Vec and GloVe word embedding methods have been used in different NLP methods in vector representation, the BERT method has been used in much less research in the literature (Khatua et al., 2019; Biswas & De, 2022; Dharma et al., 2022; Al-Aidaroos & Bamzahem, 2023; Sitender et al., 2023; Rakshit & Sarkar, 2025). It is thought that the effects of the BERT method in Twitter text analysis, which is potentially suitable for research, will provide support as an important contribution to the literature.

3.1. Material

In this study, we used a dataset containing important information about disasters, such as ID, keyword, location and actual tweet (Addison Howard et al., 2019). It also has tags in the tweet information of the corresponding dataset that identify whether the tweet contains information related to a disaster (Balakrishnan et al., 2022). The dataset contains 7613 training data, 3271 disaster tweets, 21940 unique words, 12.5 average tweet length, 29 maximum tweet length, and 1 minimum tweet length (Deb & Chanda, 2022). At the same time, the dataset consists of 7613 training data and 3263 test training data. The labelling in the dataset is reported to be manually labeled as 1 if it is a catastrophic dataset and 0 otherwise (Deb & Chanda, 2022). The dataset used in the article was obtained without changing the name of the Twitter platform and is a large-scale dataset used in Kaggle competitions.

In the study, after reading the dataset, some basic pre-processing steps were applied to bring the texts to the same standard and to achieve high performance. These steps are presented below, respectively.

- Stop words in the intended texts have been removed.
- Additional spaces, non-alphanumeric characters, internet links, symbols like @, and numeric characters that have no effect on sentiment analysis have been removed.
- Abbreviations were identified and converted to their original formats.
- After performing all the above steps, all normalized texts are converted to lower case.

3.2. BERT

A neural network model that produces different vectors for the same word is called BERT (Devlin, 2018). The BERT model is a contextual model that generates a representation for each word depending on the other words in the processed text (Deb & Chanda, 2022). It is a transformer-based model with multiple attention mechanisms and encoders. BERT architectures are similar to original transformer models in that they can learn contextual embeddings (Vaswani, 2017). Since the BERT model is transformer-based, it has encoder and decoder blocks with softmax activation function to normalize the output probabilities. The words given as input to the model are passed through a spatial encoder that assigns vectors according to their positions. As a result, it is aimed to obtain the contextual meaning of the texts given as input. It also includes multi-head attention mechanisms to determine the relationship of each word with other words in the input texts. The output from the attention vectors is transferred from the network to the decoder block. The decoders have a spatial encoder and multiheaded attention mechanisms. BERT is a computationally expensive model with monolingual classification, requiring limitations on input sentence lengths (Acheampong et al., 2021).

3.3. GloVe

GloVe is a word embedding method that combines two different words: vector and global. GloVe is also a regression model that combines the advantages of windowing models that can capture local context features and global factorization of text (Lin & Pomerleano, 2011; Cho et al., 2013). GloVe obtains a vector space with meaningful information by training on non-zero cells to obtain contextual information on the matrices obtained after digitizing the texts. GloVe is an unsupervised counting-based learning model that obtains embedding vectors by collecting word occurrence statistics (Pennington et al., 2014). In the first step, it separates the words in the text and creates word pairs. GloVE transforms the text collection that is next to each other in a certain order into a feature matrix. This transformation is done with stochastic gradient descent.

$$R = P * Q \approx R' \quad (1)$$

R in Equation 1 represents the community matrix in a certain order. The P and Q parameters consist of random values. The R' value is obtained by multiplying the P and Q values. It is important to minimize the difference between the R' matrix and the R value. It is necessary to determine what P and Q values will minimize the specified difference. The point obtained after convergence of the error with stochastic gradient descent after many iterations is called GloVe word embedding (Mikolov et al., 2013). As a result, it activates the attention mechanism by de-emphasizing rare word pairs (Pennington et al., 2014).

3.4. Word2Vec

Word2Vec is one of the word embedding methods that allows the words in the text to be digitized and represented by vectors (Mikolov et al., 2013). This method includes a single input and hidden and output layers. Different models, namely CBOW and Skip Gram, are used in literature as publications. With CBOW, the central word is predicted from the words around the center. In the Skip Gram technique, the exact opposite of the CBOW method is to predict the words around the central word from the central word. Since the Word2Vec word embedding technique can process trillions of words, it can be used in many tasks in text analysis. Word2Vec can reduce the effect of high-frequency and important words by processing rare words (Rakshit & Sarkar, 2025).

3.5. CNN

CNN architecture is used in NLP problems because of its ability to meaningfully capture local information from a word in the text. In CNN, convolution layers can be used in different numbers, different sizes or kernel values to obtain local features. Feature maps can be obtained by using different number of filters on the input text values. The resulting feature maps have various information extracted from a sentence. In the proposed CBLTwitter model, pre-processed texts are digitized with word embedding methods and then their features are obtained with convolution layers. These features are then structured by pooling layers to select the ones with high discriminative power.

3.6. BiLSTM

The LSTM model has the ability to remember more important information than classical recurrent neural network models. It solves the problem of vanishing gradients in classical recurrent neural networks thanks to the gating mechanism. Input, output and forget gates help to retain contextual information in texts for a long time. These gates consist of sigmoid layers that decide how long each input of information from the texts should pass through a gate. As the sentence lengths in texts increase, LSTM layers are insufficient to protect information (Rajesh & Hiwarkar, 2023). Since normal LSTM only allows forward propagation of information, BiLSTM architecture is proposed to extract contextual information before and after a sentence.

$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i) \quad (2)$$

$$e_t = \tau(W_c.[h_{t-1}, x_t] + b_c) \quad (3)$$

$$c_t = f_t.e_{t-1} + i_t.e_t \quad (4)$$

$$o_t = \sigma(W_o.[h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t.\tau(c_t) \quad (6)$$

$$f_t = \sigma\left(W_f.[h_{t-1}, x_t] + b_f\right) \quad (7)$$

$$i_t = \sigma(W_i.[h_{t+1}, x_t] + b_i) \quad (8)$$

$$e_t = \tau(W_c.[h_{t+1}, x_t] + b_c) \quad (9)$$

$$c_t = f_t.e_{t+1} + i_t.e_t \quad (10)$$

$$o_t = \sigma(W_o.[h_{t+1}, x_t] + b_o) \quad (11)$$

$$h_t = o_t.\tau(c_t) \quad (12)$$

$$f_t = \sigma\left(W_f.[h_{t+1}, x_t] + b_f\right) \quad (13)$$

Equations 2 through 13 represent the forward and backward movement of the BiLSTM architecture. In the relevant equations i, f, o denote input, forget and output gates respectively. x_t, h_t, c_t, b, W denote the input state, hidden state, current state, bias value, and weight matrix at time t, respectively.

3.7. Attention layer

In text processing, not every word has the same impact on the sentence meaning. Therefore, it is predicted that emphasizing the words with high disaster content with the attention layer can improve the classification performance. In the proposed CBLTwitter model, the outputs from the BiLSTM model are passed through the attention layer to assign a score to them. The outputs from the BiLSTM model are multiplied by the latent state weight values of the attention layer and given as input to the batch normalization layer where inter-layer normalization is performed. The outputs from the BiLSTM model are assigned a weight to extract the important parts of the relevant sentence. The attention layer of Sadr and Soleimandarabi (Sadr & Nazari, 2022) was implemented with modifications.

$$U_i = tanh(X_i W + b) \tag{14}$$

$$a_i = softmax(U_i u) \tag{15}$$

$$\bar{X}_i = a_i \circ X_i \tag{16}$$

U_i in Equation 14 is a context vector. With the input $u \in R^{n-h_i+1x1}$ and U_i in Equation 15, the output $a_i \in R^{n-h_i+1x1}$ is obtained from the softmax activation function. The value u in Equation 15 is the symbol defined to highlight disaster words in the texts (Sermanet et al., 2013; Sukhbaatar et al., 2015). A different representation of X_i in Equation 16 yields $\bar{X}_i$.

3.8. Proposed CBLTwitter

The model proposed in this paper is presented in Figure 1. The proposed model starts its operations by processing the raw data it receives as input. Before passing the raw input data set through the proposed CBLTwitter model, the data is pre-processed, cleaned, emojis are removed, lowercase and uppercase letters are brought to the same standardization. After these processes, an adjustment was made by using the tokenizer class to take up to 5000 words in the texts. Afterwards, the texts were converted into sequences. This step represents the first layer of the proposed CBLTwitter model. After conversion into sequences, GloVe, Word2Vec and BERT methods were used in separate experimental studies.

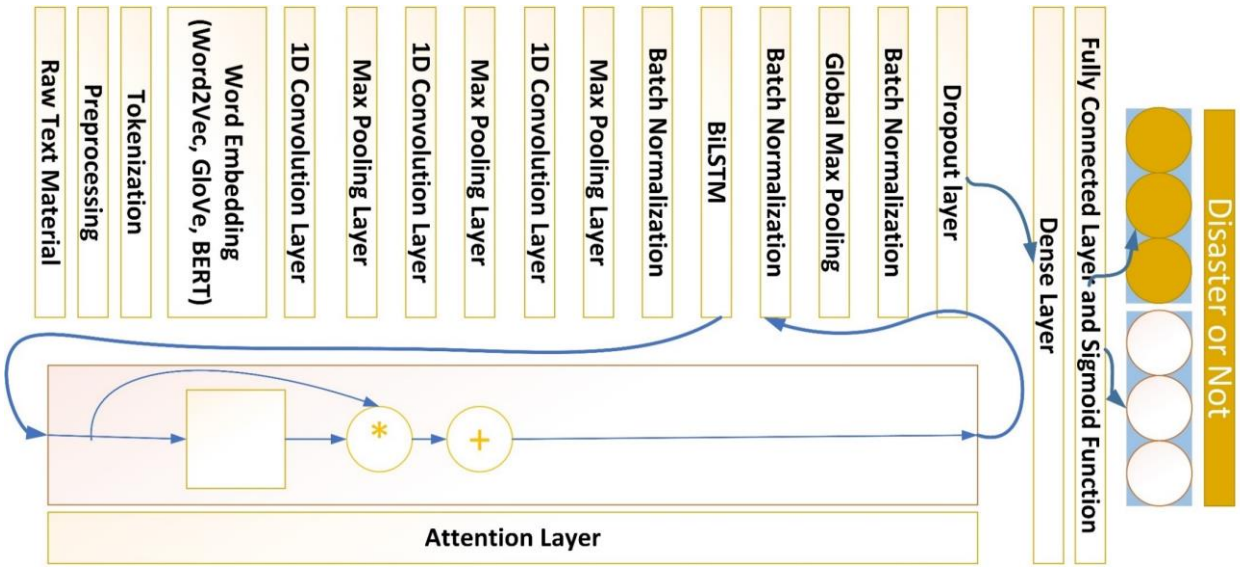

Figure 1. Proposed CBLTwitter model for analysis of disaster tweets

In the second layer of the proposed CBLTwitter model, GloVe, Word2Vec and BERT models are presented separately as input. In the third layer, a 1D convolution layer with ReLU activation function is defined as 4 kernel sizes with 128 filters. In the fourth layer, a max pooling layer with 3x3 windows is applied. In the fifth layer, a 1D convolution layer with the properties defined in the third layer is defined. In the sixth layer, max pooling is applied with the same properties as in the fourth layer. In the seventh layer, similar to the fifth layer, feature assignment was made. In the eighth layer, a max pooling layer is defined similar to the sixth layer. In the ninth layer, Batch Normalization layer is applied which performs normalization between layers. In the tenth layer, a 128-cell BiLSTM layer with high memory that can navigate between forward and backward features is defined. In the eleventh layer, attention layer is defined which assigns weights to the importance value of the input values coming from the BiLSTM layer. In the twelfth layer, Batch Normalization is defined which has the features of the ninth layer. In the thirteenth layer, the global max pooling layer is applied. In the fourteenth layer, Batch Normalization similar to the twelfth layer was added. In the fifteenth layer, a dropout layer with a dropout rate of 0.6 neurons was added to prevent the model from memorizing. In the sixteenth layer, a dense layer with 100 cells was added. The seventeenth layer is a fully connected layer that transfers the features from the previous layers to the classification layer. In the last layer, a layer with a sigmoid activation function was applied due to the two-class nature of the dataset.

The proposed model has 0.42 million parameters, while the computational complexity is 0.40 Floating Point Operations (FLOPs). When GloVe and Word2Vec methods are used, the proposed model reaches approximately 15.42 million parameters. In terms of computational complexity, the proposed model approaches approximately 0.40 million FLOPs when the GloVe and Word2Vec methods are used in order. The BERT model is quite complex compared to both other word embedding methods and the model.

$$O(BERT) = Layer\ count.\ [O(n^2.d) + O(n.d^2)] \quad (17)$$

The complexity of BERT is calculated according to Equation 17. The 12-layered version is used in the BERT word embedding method. In Equation 17, n indicates the length of the input sequence, while d indicates the hidden size. The d value is 768 in BERT base operations. n is taken as 500 in this study. In this case, it will have over a million parameters obtained using BERT. In the BERT base method, approximately 110 million parameters and approximately 1.86 Giga Floating Point Operations (GFLOPs) value is reached in terms of computational complexity.

4. Results and discussion

A proposed CBLTwitter model is presented to create a high performance system that focuses on important words after preprocessing Disaster texts using NLP techniques. Experimental studies were carried out separately with each of the proposed CBLTwitter model, GloVe, Word2Vec and BERT models. The experimental studies were carried out on a computer with NVIDIA GeForce RTX 3060 graphics card and AMD Ryzen 7 5800H processor. Spyder development environment was preferred as the working environment. Although there are experimental environments such as Google Colab, the Spyder environment is less costly and more useful for processes that require long training times such as BERT. For the reasons mentioned above, the algorithm was developed on a local computer environment.

In the experimental studies, the proposed CBLTwitter model was trained based on the parameters presented in Table 1. The parameters were determined based on the reference articles used in the creation of the article. When the studies in the literature were analysed in detail, the parameters in Table 1 were assigned as these parameters were predicted to give superior performance results. The main reason for using these metrics is to make comparisons with recent studies in the literature.

Table 1. Proposed CBLTwitter hyperparameter values according to word embedding methods

Number	Word embedding types	Hyper parameters	Values
1	BERT, Glove and Word2Vec	Optimizer	Adam
2	Glove and Word2Vec	Learning rate	0.001
3	Glove and Word2Vec	Number of epochs	30
4	Glove and Word2Vec	Loss functions	Binary cross entropy

Looking at the results given in Table 2, the BERT transformer word embedding method outperformed other word embedders by contributing positively to the proposed CBLTwitter model, which is composed of the right preprocessing and the right combination of layers. One of the underlying reasons behind the competitive results of the proposed CBLTwitter model with the studies in the literature lies in the correct analysis of the studies in the literature before the study and the combination of important methods in the correct order. Precision, Recall, F1 score and Accuracy formulas were calculated using common formulas from the study in (Çetiner, 2022).

Table 2. Performance result of the proposed CBLTwitter model on different word embeddings

Word embedding	Type	Precision	Recall	F1 score	Accuracy
GloVe	Train	0.92	0.93	0.92	0.93
GloVe	Validation	0.91	0.92	0.91	0.92
Word2Vec	Train	0.94	0.93	0.93	0.94
Word2Vec	Validation	0.93	0.92	0.92	0.93
BERT	Train	**0.98**	**0.97**	**0.98**	**0.97**
BERT	Validation	**0.98**	0.95	0.96	0.96

Among the performance results given in Table 2, the highest results were obtained with the BERT word embedder. The training and validation results obtained with BERT are presented in Figures. 2-5. According to the accuracy performance results given in Figure 2, the training accuracy is 0.97 while the validation performance result is 0.96. Figure 3 shows the precision performance results of the proposed CBLTwitter model by changing only the word embedder while keeping everything else the same. According to the obtained precision results, the training and validation precision value reached 0.98. Figure 4 shows the recall performance results obtained with the same conditions as in the previous two figures. According to the recall results, the training and validation precision values are 0.97 and 0.95 respectively.

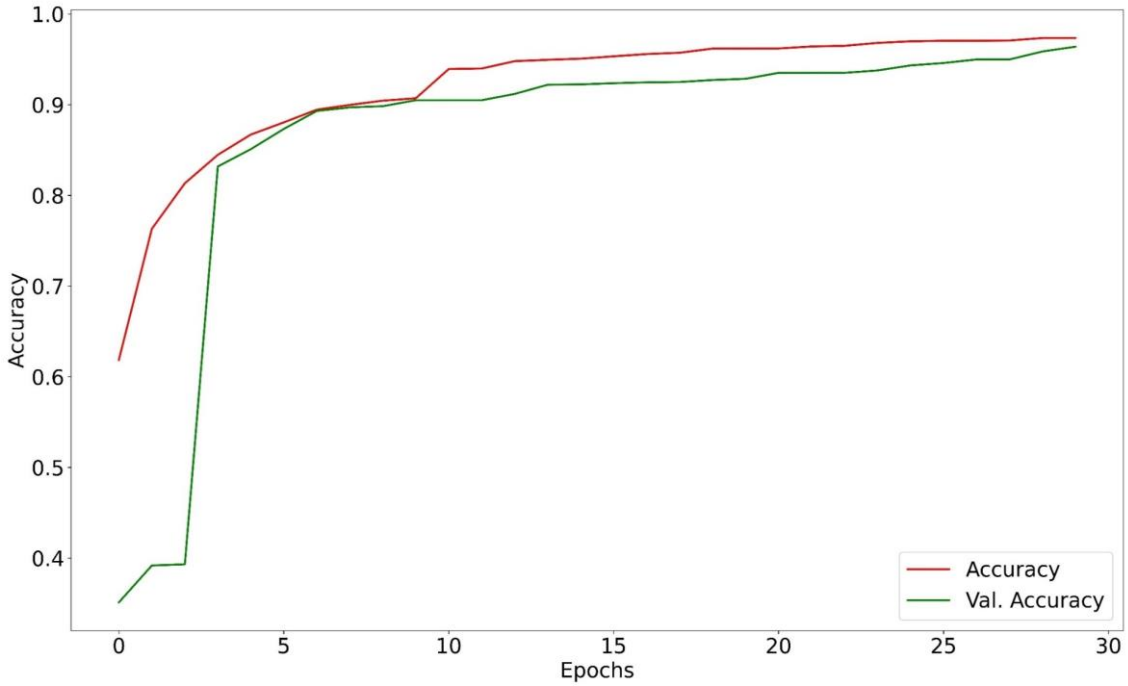

Figure 2. Accuracy performance result of the proposed CBLTwitter model using BERT word embedder

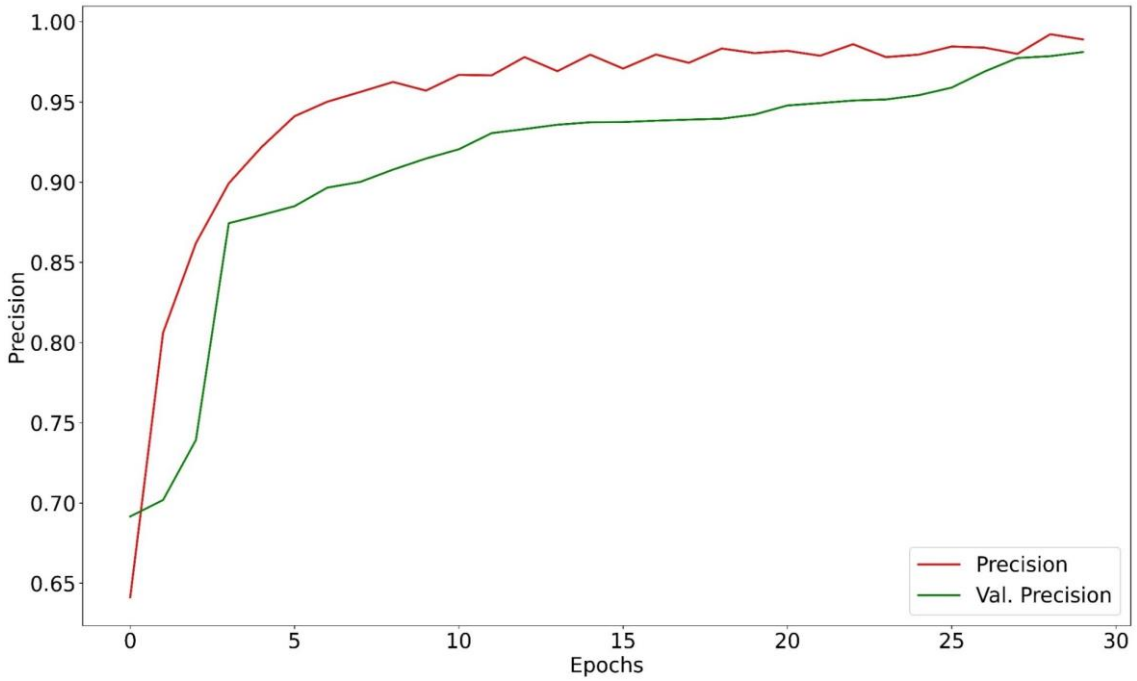

Figure 3. Precision performance result of the proposed CBLTwitter model using BERT word embedder

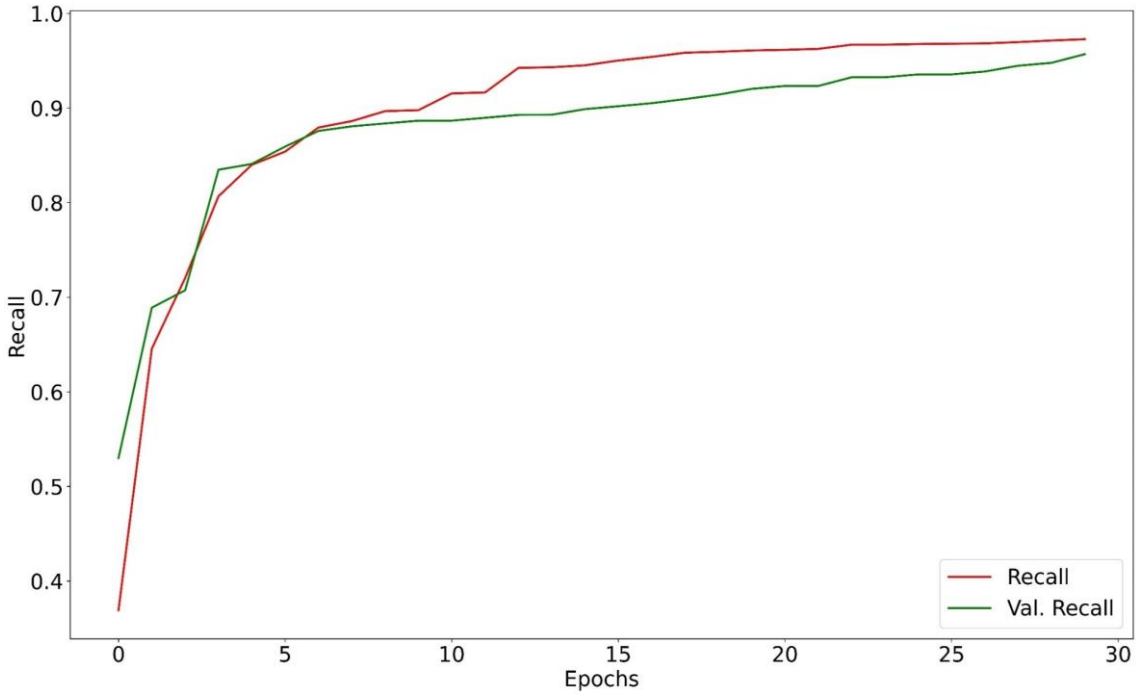

Figure 4. Recall performance result of the proposed CBLTwitter model using BERT word embedder

The visual results given in Figure 5 provide a clearer representation of the numerical figures given in Table 2. In Figure 5, the F1 score training and validation performance result of the proposed CBLTwitter model using the BERT word embedder with the same conditions as in Figure 2, Figure 3 and Figure 4 reaches 0.98 and 0.96 respectively.

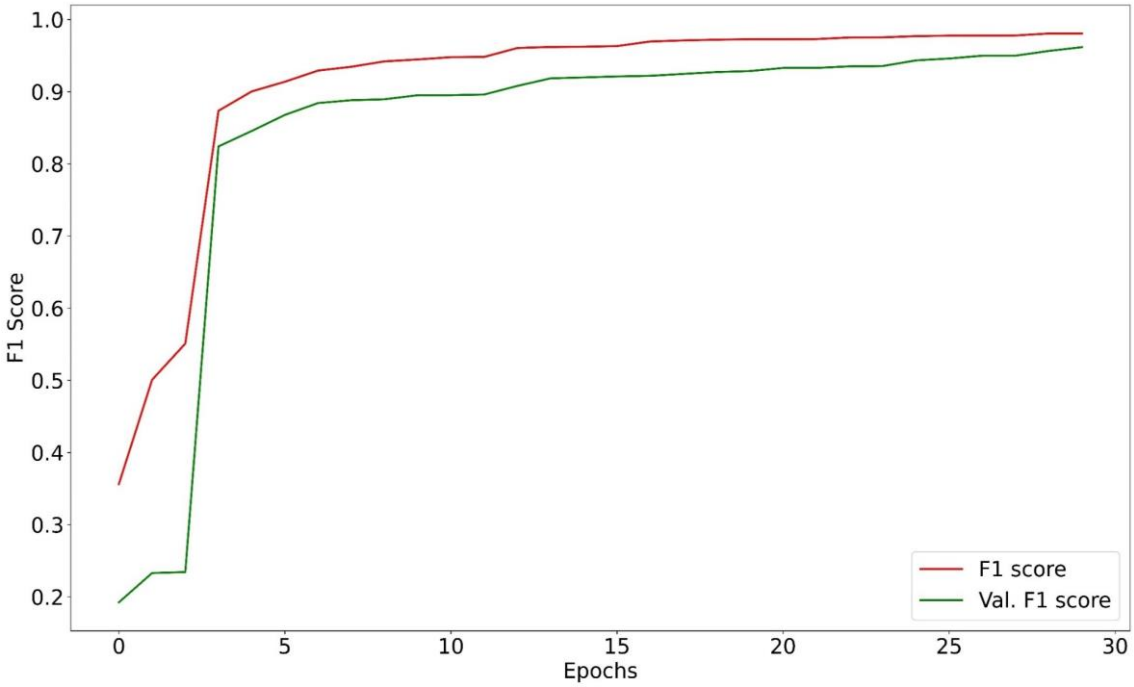

Figure 5. F1 score performance result of the proposed CBLTwitter model using BERT word embedder

By looking at the performance results given in Figures 2, 3, 4 and 5, it is possible to draw a conclusion about the proposed CBLTwitter model. However, in order to provide a more reliable result, a comparison of the proposed CBLTwitter model with state-of-art (SOTA) studies that have been conducted in recent years has been performed. The proposed CBLTwitter model is competitive with studies in the literature using the same dataset. The studies (Sukhbaatar et al., 2015), (Song & Huang, 2021), (Manthena, 2023), (R. et al., 2023), and (Mahajan et al., 2024) in the literature are the studies prepared using the dataset used in this paper. A comparison was made with these studies, each of which is more valuable than the other, and a conclusion was reached. Table 3 shows the performance results obtained with machine learning techniques as well as the performance results obtained with recurrent neural networks. There are also results obtained with convolutional-based deep learning networks.

Table 3. State-of-art (SOTA) performance comparisons on the same dataset

Model & ref.	Precision	Recall	F1 score	Accuracy
BERT-LSTM (Sukhbaatar et al., 2015)	0.85	0.84	0.84	0.87
Logistic Regression (R. et al., 2023)	-	-	-	0.50
BernoulliNB (R. et al., 2023)	-	-	-	0.80
SGD (R. et al., 2023)	-	-	-	0.78
XGB (R. et al., 2023)	-	-	-	0.80
Random Forest (R. et al., 2023)	-	-	-	0.78
BERT (Mahajan et al., 2024)	0.89	0.90	0.90	0.90
TF-IDF (Mahajan et al., 2024)	0.85	0.64	0.73	0.80
BiLSTM-GloVe (Manthena, 2023)	0.80	0.71	0.75	0.80
LSTM-GloVe (Manthena, 2023)	0.75	0.68	0.72	0.77
Word2Vec-BiLSTM-CNN (Song & Huang, 2021)	0.88	0.87	0.87	-
BiLSTM (Song & Huang, 2021)	0.85	0.84	0.84	-
BiLSTM-CNN (Song & Huang, 2021)	0.86	0.85	0.85	-
Proposed CBLTwitter with BERT	**0.98**	**0.95**	**0.96**	**0.96**
Proposed CBLTwitter with Word2Vec	0.93	0.92	0.92	0.93
Proposed CBLTwitter with GloVe	0.91	0.92	0.91	0.92

When all the studies in Table 3 are examined in detail, it can be seen that preprocessing algorithms are important in natural language processing problems. In addition, it can be said that layers and connections between layers are also very important.

What makes this study superior to other studies in the literature is the robustness of the algorithm. The strengthening of the architectures that provide local and global features in the algorithm with the attention mechanism that gives more weight to important words has increased the effectiveness of the proposed CBLTwitter model. The test accuracy results of the models proposed in the study, namely proposed CBLTwitter with BERT, proposed CBLTwitter with Word2Vec, and proposed CBLTwitter with GloVe, are 96.71%, 93.58%, and 92.48%, respectively.

The performance results obtained using the attention layer of the proposed CBLTwitter model are given in Table 3. A serious performance loss was experienced in the ablation experiments conducted to determine the effect of the attention layer, which is one of the layers of the model. By using BERT word embedding method, precision, recall, F1 score and accuracy performance values decrease to 0.92, 0.89, 0.90, 0.90 respectively. In ablation tests obtained by using Word2Vec word embedding method without attention layer, precision, recall, F1 score and accuracy performance values decrease to 0.85, 0.86, 0.86, 0.87 respectively. Finally, in the ablation tests performed with the GloVe word embedding layer without the attention layer, the precision, recall, F1 score and accuracy performance values decrease to 0.86, 0.85, 0.86 and 0.87, respectively. When the performance results obtained are evaluated collectively, a serious decrease is observed in all performance criteria. In this case, since it is not possible to compete with some studies in the literature, it can be said that the attention layer increases the classification performance by highlighting the words containing high disaster.

Table 4. State-of-art (SOTA) performance comparisons on the different datasets

Model & ref.	Precision	Recall	F1 score	Accuracy	Dataset
Proposed CBLTwitter with BERT	**0.95**	**0.96**	**0.95**	**0.97**	Tweets Airline (Eight, 2019)
Proposed CBLTwitter with Word2Vec	0.94	0.95	0.95	0.96	Tweets Airline (Eight, 2019)
Proposed CBLTwitter with GloVe	0.95	0.96	0.95	0.95	Tweets Airline (Eight, 2019)
Proposed CBLTwitter with BERT	**0.96**	**0.96**	**0.96**	**0.97**	Twitter SemEval (Rosenthal et al., 2017)
Proposed CBLTwitter with Word2Vec	0.93	0.92	0.92	0.93	Twitter SemEval (Rosenthal et al., 2017)
Proposed CBLTwitter with GloVe	0.91	0.92	0.91	0.92	Twitter SemEval (Rosenthal et al., 2017)

In order to improve the reliability of the study, the proposed model was tested on two different datasets at this stage. Table 4 shares the performance results obtained on two different datasets used for benchmarking purposes in the literature. The names of the datasets used to test the reliability of the study are Tweets Airline and Twitter SemEval datasets. One of these datasets, Tweets Airline, is a dataset with 14,460 user opinions obtained as a result of collecting data from six different airlines (Eight, 2019). The other dataset, Tweets SemEval, is a dataset prepared to support sentiment analysis studies on Twitter data (Rosenthal et al., 2017). Although there are different types of SemEval dataset, a structure consisting of 17,750 tweets was used in the tests conducted to test the reliability of the proposed model in the study. The performance results obtained are presented in Table 4. When the shared performance results are examined, it is seen that the proposed CBLTwitter model achieves results at a level that can compete with the studies in the literature in different datasets.

5. Conclusions

In training the data used in this study, BERT, a transformer-based neural network model, consumed a significant amount of memory. Training a dataset with a large amount of tweet data with a BERT model, which requires high computational cost and cloud storage capacity, also requires a lot of time. In addition to the performance results obtained with BERT, performance results were obtained with GloVe and Word2Vec methods by changing the word embedder in the proposed CBLTwitter model. Considering the performance results obtained, it can be said that all three results are competitive with the studies in the literature using the same dataset. Although BERT, which is one of the state-of-the-art structures, has been used in many NLP problems, it is difficult to adapt to every problem without awareness because it is pre-trained. Considering that the dataset used in the study is a large dataset, it may seem normal that the training time is long. However, when the training times of GloVe and Word2Vec word embedding methods other than BERT are compared on the same dataset, it is seen that the BERT model has a very high training time. The training time of the model with 30 epochs of iteration on the same dataset is six and five minutes for GloVe and Word2Vec word embedding algorithms, respectively. However, the training time of the BERT model with 30 epochs of iteration on the same dataset takes approximately thirty-six hours. Since the training times specified are highly dependent on hardware resources, they have not been evaluated much. However, it can be stated that word

embedding methods are important in terms of their effects on the training time of the proposed CBLTwitter model. In addition, considering the closeness of the performance of GloVe and Word2Vec methods to the BERT model, GloVe and Word2Vec methods can be preferred instead of BERT in terms of speed and performance.

Considering the aforementioned situation, the BERT model is one of the models with a very long training time in natural language processing problems. In future studies, it is aimed to develop word embedding methods that provide high performance results with less computational cost despite the encoder and decoder structure.

Author contribution

The corresponding author carried out the coding, writing and editing of the experimental studies. The other author carried out the literature review.

Declaration of ethical code

The authors declare that this study does not require ethical committee approval or any legal permission.

Conflicts of interest

The authors declare no competing interests.

References

Acheampong, F. A., Nunoo-Mensah, H., & Chen, W. (2021). Transformer models for text-based emotion detection: a review of BERT-based approaches. *Artificial Intelligence Review*, *54*(8), 5789–5829.

Addison Howard, devrishi, Phil Culliton, Y. G. (2019, December 20). *Natural language processing with disaster tweets*. https://kaggle.com/competitions/nlp-getting-started/data.

Al-Aidaroos, A. S., & Bamzahem, S. (2023). The impact of GloVe and Word2Vec word-embedding technologies on bug localization with convolutional neural network. *International Journal of Science and Engineering Applications,12*(1), 108-111.

Alami, S., & Elbeqqali, O. (2015). Cybercrime profiling: text mining techniques to detect and predict criminal activities in microblog posts. *2015 10th International Conference on Intelligent Systems: Theories and Applications (SITA)* (pp. 1–5), Rabat.

Balakrishnan, V., Shi, Z., Law, C. L., Lim, R., Teh, L. L., Fan, Y., & Periasamy, J. (2022). A comprehensive analysis of transformer-deep neural network models in twitter disaster detection. In *Mathematics, 10*(24), 4664.

Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: approaches, challenges and trends. *Knowledge-Based Systems*, *226*, 107134.

Biswas, R., & De, S. (2022). A comparative study on improving word embeddings beyond Word2Vec and GloVe. *2022 Seventh International Conference on Parallel, Distributed and Grid Computing (PDGC)*, (pp. 113–118), Solan.

Çetiner, H. (2022). Multi-label text analysis with a CNN and LSTM based hybrid deep learning model. *Adıyaman Üniversitesi Mühendislik Bilimleri Dergisi*, *9*(17), 447-457.

Çetiner, H. (2023). Cataract disease classification from fundus images with transfer learning based deep learning model on two ocular disease datasets. *Gümüşhane Üniversitesi Fen Bilimleri Dergisi*, *13*(2), 258-269.

Çetiner, H. (2024). Fake news detection and classification with recurrent neural network based deep learning approaches. *Osmaniye Korkut Ata Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, *7*(3), 973–993.

Çetiner, M. (2022). *Sürdürülebilir moda ürünlerinin derin öğrenme yaklaşımı kullanarak analizi* [Doktora Tezi, Süleyman Demirel Üniversitesi Sosyal Bilimler Enstitüsü].

Cho, H. C., Okazaki, N., & Inui, K. (2013). Inducing context gazetteers from encyclopedic databases for named entity

recognition. In *Advances in Knowledge Discovery and Data Mining: 17th Pacific-Asia Conference, PAKDD 2013, Gold Coast, Australia, April 14-17, 2013, Proceedings, Part I 17* (pp. 378-389). Springer.

Deb, S., & Chanda, A. K. (2022). Comparative analysis of contextual and context-free embeddings in disaster prediction from Twitter data. *Machine Learning with Applications, 7*, 100253.

Devlin, J. (2018). Bert: pre-training of deep bidirectional transformers for language understanding. *ArXiv Preprint ArXiv:1810.04805*.

Dharma, E. M., Gaol, F. L., Warnars, H., & Soewito, B. (2022). The accuracy comparison among word2vec, glove, and fasttext towards convolution neural network (CNN) text classification. *J Theor Appl Inf Technol*, *100*(2), 31.

Eight, F. (2019, February 21). *Twitter Airline Sentiment*. https://www.kaggle.com/datasets/crowdflower/twitter-airline-sentiment.

Feng, X., Angkawisittpan, N., & Yang, X. (2024). A CNN-BiLSTM algorithm for weibo emotion classification with attention mechanism. *Mathematical Models in Engineering*, *10*(2), 87-97.

Huang, Q., Chen, R., Zheng, X., & Dong, Z. (2017). Deep sentiment representation based on CNN and LSTM. *2017 International Conference on Green Informatics (ICGI)* (pp. 30–33), Fuzhou.

Khan, S., & Yairi, T. (2018). A review on the application of deep learning in system health management. *Mechanical Systems and Signal Processing*, *107*, 241–265.

Khatua, A., Khatua, A., & Cambria, E. (2019). A tale of two epidemics: contextual Word2Vec for classifying twitter streams during outbreaks. *Information Processing & Management*, *56*(1), 247–257.

Kim, S., & Lee, S. P. (2023). A BiLSTM–transformer and 2D CNN architecture for emotion recognition from speech. *Electronics, 12*(19), 4034.

Kishwar, A., & Zafar, A. (2023). Fake news detection on Pakistani news using machine learning and deep learning. *Expert Systems with Applications*, *211*, 118558.

Kowsher, M., Tahabilder, A., Islam Sanjid, M. Z., Prottasha, N. J., Uddin, M. S., Hossain, M. A., & Kader Jilani, M. A. (2021). LSTM-ANN & BiLSTM-ANN: Hybrid deep learning models for enhanced classification accuracy. *Procedia Computer Science*, *193*, 131–140.

Lin, K., & Pomerleano, D. (2011). Global matrix factorizations. *Mathematical Research Letters*, *20*.

Mahajan, P., Raghuwanshi, P., Setia, H., & Randhawa, P. (2024). A multi-model approach for disaster-related tweets: a comparative study of machine learning and neural network models. *Journal of Computers, Mechanical and Management*, *3*(2), 19-24.

Manthena, S. P. (2023). *Leveraging tweets for rapid disaster response using BERT-BiLSTM-CNN model* [Master of Science, San Jose State University Department of Computer Science].

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Neethu, M. S., & Rajasree, R. (2013). Sentiment analysis in twitter using machine learning techniques. *2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)* (pp. 1–5), Tiruchengode.

Pennington, J., Socher, R., & Manning, C. D. (2014, October). Glove: global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532-1543), Doha.

Priya, C. S. R., & Deepalakshmi, P. (2023). Sentiment analysis from unstructured hotel reviews data in social network using deep learning techniques. *International Journal of Information Technology*, *15*(7), 3563–3574.

R., M., S., M., OS, N., & E., T. (2023). An enhanced framework for disaster-related tweet classification using machine learning techniques. *2023 International Conference on Inventive Computation Technologies (ICICT)* (pp. 108–111), Nepal.

Rajesh, A., & Hiwarkar, T. (2023). Sentiment analysis from textual data using multiple channels deep learning models. *Journal of Electrical Systems and Information Technology*, *10*(1), 56.

Rakshit, P., & Sarkar, A. (2025). A supervised deep learning-based sentiment analysis by the implementation of Word2Vec and GloVe embedding techniques. *Multimedia Tools and Applications*, *84*, 979-1012.

Rosenthal, S., Farra, N., & Nakov, P. (2017). SemEval-2017 task 4: Sentiment analysis in twitter. In S. Bethard, M. Carpuat, M. Apidianaki, S. M. Mohammad, D. Cer, D. Jurgens (eds.), *Proceedings of the 11th international workshop on semantic evaluation (SemEval-2017)* (pp. 502–518). Vancouver: Association for Computational Linguistics.

Sadr, H., & Nazari Soleimandarabi, M. (2022). A CNN-TL: attention-based convolutional neural network coupling with transfer learning and contextualized word representation for enhancing the performance of sentiment classification. *The Journal of Supercomputing*, *78*(7), 10149–10175.

Semary, N. A., Ahmed, W., Amin, K., Pławiak, P., & Hammad, M. (2023). Improving sentiment classification using a RoBERTa-based hybrid model. *Frontiers in Human Neuroscience*, *17*.

Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., & LeCun, Y. (2013). Overfeat: integrated recognition, localization and detection using convolutional networks. *ArXiv Preprint ArXiv:1312.6229*.

Shaik, T., Tao, X., Dann, C., Xie, H., Li, Y., & Galligan, L. (2023). Sentiment analysis and opinion mining on educational data: a survey. *Natural Language Processing Journal*, *2*, 100003.

Sitaula, C., & Shahi, T. B. (2024). Multi-channel CNN to classify nepali COVID-19 related tweets using hybrid features. *Journal of Ambient Intelligence and Humanized Computing*, *15*(3), 2047–2056.

Sitender, Sangeeta, Sushma, N. S., & Sharma, S. K. (2023). Effect of GloVe, Word2Vec and FastText embedding on english and hindi neural machine translation systems. In *Proceedings of Data Analytics and Management: ICDAM 2022* (pp. 433-447). Singapore: Springer Nature Singapore.

Song, G., & Huang, D. (2021). A sentiment-aware contextual model for real-time disaster prediction using twitter data. *Future Internet*, *13*(7), 163.

Sukhbaatar, S., Weston, J., & Fergus, R. (2015). End-to-end memory networks. *Advances in Neural Information Processing Systems*, *28*.

Tam, S., Said, R. Ben, & Tanriöver, Ö. Ö. (2021). A ConvBiLSTM deep learning model-based approach for twitter sentiment classification. *IEEE Access*, *9*, 41283–41293.

Tan, K. L., Lee, C. P., Anbananthen, K. S. M., & Lim, K. M. (2022). RoBERTa-LSTM: a hybrid model for sentiment analysis with transformer and recurrent neural network. *IEEE Access*, *10*, 21517–21525.

Vadivukarassi, M., Puviarasan, N., & Aruna, P. (2018). An exploration of airline sentimental tweets with different classification model. *International Journal for Research in Engineering Application & Management*, *4*(2).

Vaswani, A. (2017). Attention is all you need. *ArXiv Preprint ArXiv:1706.03762*.

Wankhade, M., Annavarapu, C. S. R., & Abraham, A. (2024). CBMAFM: CNN-BiLSTM multi-attention fusion mechanism for sentiment classification. *Multimedia Tools and Applications*, *83*(17), 51755–51786.

Yang, Y., & Li, S. (2024). Entity overlapping relation extracting algorithm based on CNN and BERT. *IEEE Access*, 1.

Yeboah, P. N., & Baz Musah, H. B. (2022). NLP technique for malware detection using 1D CNN fusion model. *Security and Communication Networks*, *2022*(1), 2957203.

Zhao, J., Liu, K., & Xu, L. (2016). *Sentiment analysis: mining opinions, sentiments, and emotions*. Cambridge University Press.