International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2020): 7.803

A Russia-Ukraine Conflict Tweets Sentiment Analysis Using Bidirectional LSTM Network

Eman Sedqy Ibraihm Shlkamy[1], Khaled Mohammed Mahar[2], Ahmed Ahmed Hesham Sedky[3]

[1] Arab Academy for Science Technology and Maritime Transport, College of Computing and Information Technology, Information Systems Department, Alexandria Egypt
E.Shlkamy15020[at]student.aast.edu

[2]Arab Academy for Science Technology and Maritime Transport, College of Computing and Information Technology, Information Systems Department, Alexandria Egypt
khmahar[at]aast.edu

[3]Arab Academy for Science Technology and Maritime Transport, College of Computing and Information Technology, Information Systems Department, Alexandria Egypt
ahsedkyy[at]aast.edu

Abstract: ***Sentiment analysis techniques have a vital role in analyzing people's opinions. The continuous and rapid growth of data posted on social media sites is the fuel that draws people's opinions. Despite the fact that the vast majority of research focuses on analyzing sentiment to study the impact of the war on the global economy. Furthermore, the actions of national leaders or other powerful figures have typically received more attention in the research of international conflict than public emotions and opinions. The purpose of this paper is to go over some of the most relevant works on sentiment analysis, which are limited by a simple architecture and focus on analyzing public emotions and opinions during the Russia-Ukraine Conflict as compared to leaders and other powerful figures. This paper proposes to use a single bidirectional LSTM network for English tweet sentiment analysis using a classification of positive, negative, and neutral as a multi-class classification approach. We utilized one bidirectional long short-term memory (Bi-LSTM) layer along with the global Max pooling ID mechanism and achieved an accuracy of 91.79%. The results of the proposed framework show well performance over previous studies with complex structures that have previously been proposed. Measuring performance in terms of accuracy. During crises, it is crucial to pay attention to simple architectural models to solve similar problems without complex architectural neural networks.***

Keywords: Sentiment Analysis, Bidirectional LSTM, Deep learning, Natural language processing (NLP)

1. Introduction

Humans have emotions as part of their nature. As a result, they help us gain a better understanding of what we are experiencing and help us react appropriately. Understanding emotions is the key to understanding human behavior. We are capable of identifying our feelings and putting them into words. With practice and time, we become better at identifying our feelings and recognizing them. When you are aware of your emotions, it is easier to decide what you need and want. This is because being emotionally aware can help us communicate our feelings more effectively, prevent or resolve conflicts, and move through difficult emotions more easily [1].

Neurosurgeon Antonio R. and his colleague have produced solid evidence that emotions and decision-making are essentially intertwined. However, Damasio has shown that without emotions, we might not make the right decisions [2].

During a conflict lack of social bonds contribute to negative feelings like anxiety other mental health outcomes, including depression, loneliness, and social anxiety are greatly reduced when negative emotions are eliminated [3].

Human activity is largely driven by opinions because they significantly impact how we behave. When we have to make a decision. According to [4] Sentiment and emotions are strongly intertwined.

Emotions are a complex group of neurological expressions comprising three components: sentimental occurrences, psychological reactions, and sociological reactions [5]. Early recognition of the problem is of utmost importance.

An essential skill for the imitation of human intelligence, understanding emotions is one of the most crucial components of personal growth and development. Emotion processing is crucial for the development of AI, as well as for the closely related problem of polarity identification.

Nearly all researchers in Natural Language Processing and Computational Linguistics assume that speakers have some emotion value or sentiment regarding certain features of a topic. This has given rise to the developing disciplines of affective computing and sentiment analysis, which make use of information retrieval and human-computer interaction[6]

Natural Language Processing (NLP) is the study of how computers and humans interact with language. Text preprocessing software aims to reduce numerous variants of a word to just one [7]. Classifying text begins with preprocessing. They can reduce processing times, resource requirements, and accuracy significantly [8]. Natural language processing (NLP), a subfield of sentiment analysis, receives a lot of attention from researchers [9].

A growing need to process opinionated web and social network content has made sentiment analysis (SA) one of the fastest-evolving research areas. It uses layers of algorithms to create neural networks, which are artificial replicas of the brain's structure. In order to make decisions, it mimics the functions of the human brain for managing data and forming patterns. The use of deep learning in sentiment analysis has also grown in popularity in recent years. The design and operation of the human brain which is capable of self-learning served as an inspiration for deep learning algorithms. Sentiment analysis has been developed. Machine learning-based and lexicon-based systems are the two main types. Although lexicon-based and machine-learning techniques have produced results with a high degree of precision, the feature engineering required is difficult and time-consuming. Among the fastest growing areas of machine learning is deep learning, which relies on artificial neural networks (ANNs). Deep learning algorithms have mostly been applied to sentiment analysis. Considering their capacity to automatically learn and construct input models using datasets

Thus, deep learning methods make it easier to create computational models because they can learn from data sets without the need for manual attribute selection. According to a recent study on deep learning [10].

Deep networks show improved performance through surveys of deep learning methods for sentiment classification [11]. Recurrent neural networks (RNNs) are deep learning neural networks designed specifically to learn sequences of data and are mainly used for textual data classification[12] .Nevertheless, RNNs suffer from vanishing gradients when it comes to long sequences of data. LSTM neural networks were proposed as a solution to this problem and have proven to be efficient in many real-world issues [13].

Bidirectional LSTM models, LSTM model can get more semantic features, which is helpful for sentiment classification, [14] also the results show the effectiveness of Bidirectional LSTM to perform sequential data models [12]. The proposed single-layered Bidirectional LSTM-based architecture is computationally efficient and can be recommended for real-time applications in the field of sentiment analysis [15].

Recurrent neural networks have the drawback of having short-term memory that allows earlier information to be stored in the present cell. For longer sequences, however, this skill rapidly declines. LSTM models were developed to address this issue and have even longer memory retention. To achieve a more meaningful result, bidirectional long-term memory (Bi-LSTM) was introduced, combining LSTM layers from both directions.

The contribution of this paper is to provide insights into English, the world's most spoken language today[16]. Statistics in 2020 show that about 1,132 million speakers of the English language are around the world. In international business, tourism, technology, and many other areas, it has become the default language. Bilingual people who speak Spanish and English can understand 1 in 3 Internet users. Moreover, over 60% of everything published on the web can be accessed. The application uses a simple architecture and goes beyond the limitations of Sentiment Analysis of English Tweets. This shows the extraction of emotions hidden in Twitter posts during the war for public emotions and opinion. However, more attention has been paid to national leaders and other powerful figures in international conflict research than to public opinions and emotions[17].

The paper is divided into the following sections: Section 2 Related work while Section 3 gives details of the methodology used. Section 4 contains the Experiment and Results, Section 5 displays the Data Analysis, and Section 6 describes the Model Evaluation. Sections 7 are dedicated to the discussion and the achievements of our research. Finally, in section 8 we have the conclusions and future work for this paper.

2. Related Work

In this section, the literature on sentiment analysis has been presented by highlighting the deep-learning approaches used in sentiment analysis.

During the past few decades, significant research has been done on sentiment analysis as an NLP task. They compare different deep learning architectures, such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) recurrent neural networks. Additionally, different pre-trained word embedding models are studied in combination with these models. For the 3-class dataset, they found that BERT followed by a Bi-LSTM provides marginally better results than the other models, however utilizing BERT significantly raises the computational cost. It takes less time to train a CNN and the results are adequate 0.9046%. Ten class outcomes of a hybrid model composed of a bidirectional LSTM followed by a CNN provide the best results. On the other hand, the CNN model requires less training time than the bidirectional LSTM and hybrid models by 0.8944 [18].

Khan Hasib and Ahsan Habib suggested a new form of deep learning with different features to classify emotions, Deep learning methods, such as four-layer DNN and CNN, have been used for labeling, which combines to provide an accuracy of 91%[19].

Zahra Rajabi and Ozlem Uzuner have investigated several approaches to address emotion classification in short, informal (Twitter) text multi-filter CNN-Bi-LSTM outperforms existing models, achieving 85.1% accuracy [20].

Arnab Roy and Muneendra Ojha performed comparisons of tweet sentiment classification using Google BERT, attention-based Bidirectional LSTM, and Convolutional Neural Networks (CNNs). Machine learning techniques were less effective and accurate at studying emotions as compared to models [21]

According to [22] sentiment analysis on social media, the Bi-

LSTM-based training model outperformed the traditional LSTM model in terms of accuracy and F1-measure for sentiment analysis on social media.

The proposed model in [23] uses CNN and Bi-LSTM sentiment analysis for user reviews using a Bi-LSTM self-attention-based CNN model. Based on the experimental results, the model achieves a high level of classification accuracy and F1 measure value is achieved.

On the other hand, Sharat Sachin and Abha Tripathi have implemented the baseline models for LSTM, GRU, and Bi-LSTM, and Bi-GRU on an Amazon review dataset. The bidirectional gated recurrent units perform the best in terms of accuracy, with an accuracy of 71.19%. Bi-GRU model outperforms the other as it achieved a higher value against each performance measure in this work, researchers attempted a survey of different deep learning techniques that have been applied to sentiment classification and analysis. Sakirin Tam [25] proposed an integrated structure of CNN and Bi-LSTM model, which was implemented in Twitter sentiment classification with ConvBiLSTM. ConvBiLSTM model with Word2Vec on the retrieved Tweets dataset outperformed the other models with 91.13% accuracy [24].

For aspect level opinion mining, they propose and compare two interactive attention neural networks, one of which uses two Bi-directional Long-Short-Term Memory (BL-STM) and the other two convolutional neural networks (CNN). They discovered that All LSTM-based models outperformed the Majority method, LSTM can enhance performances and learn better representations for polarity classification tasks. using BLSTM to model opinion targets and context helps to learn hidden word semantics and improves represent targets and context, which contributes to aspect-level opinion polarity classification. Regarding the CANN model, it outperforms CNN and slightly improves accuracy across both datasets Laptop and Restaurant with accuracy of BANN 73.51%, 80.71% , CANN 69.75%, 78.04%. [26].

Using the IMDB movie reviews dataset, authors compared the performance of word embedding models (Word2vec and Doc2vec) and deep learning models (CNN, LSTM, GRU, and CNN-LSTM). The results show that doc2vec models and CNN perform well on sentiment classification, and it was noted the significance of combining models to improve accuracy. In the second comparison, between CNN, LSTM, GRU and CNN-LSTM models for sentiment classification, they found that when combining the CNN with LSTM, this improves the result comparing with the LSTM model alone, CNN+LSTM 86.94 % [11].

In the proposed work [27], authors developed a novel approach to improve sentence-level sentiment analysis using sentence type classification. The approach employs Bi-LSTM-CRF to extract the target expression in opinionated sentences. The experimental results of sentiment classification accuracy on multiple benchmark datasets are 82.3 %, 48.5%, 88.3% and 85.4%.

Authors provided a method for analyzing opinions in long texts that combines Convolutional Neural Networks (CNN) and Bidirectional Long Short-Term Memory (Bi-LSTM) models with Doc2vec embedding. The model was applied to French newspaper articles and fared better than other models with 90.66% accuracy. The model is compared with CNN, LSTM, and Bi-LSTM [28].

Feature Enhanced Attention CNN-Bi-LSTM (FEA-NN) is an aspect-level neural network proposed in [29]. the method consists extracting a high-level phrase representation sequence from the embedding layer using CNN, which provides effective support for subsequent encoding tasks. To improve the quality of context coding and to obtain semantic information, they used Bi-LSTM to capture local features of phrases as well as global and temporal sentence semantics, then evaluated the model on three datasets. The experimental results varied the effectiveness of the proposed neural network and showed that the model achieves performance results such as Restaurant, Laptop, and Twitter with accuracy of 83.21, %,78, 55%, and 73.31% respectively.

Long short-term memory (LSTM) and Bi-directional LSTM (Bi-LSTM) were included in the recurrent neural network (RNN) oriented architecture used to evaluate the performance of the predictive models, with LSTM obtaining an accuracy of 90.59% and Bi-LSTM achieving 90.83% [30]. The authors of [31] demonstrate how the psychology and behavior of society may lessen the impact of the current economic and social crises during the conflict between Ukraine and Russia. The approach employs two Bi-RNN with VADER a pre-trained sentiment analyzer with large dataset in around 11.15 million tweets. The experimental results of sentiment binary classification accuracy with 93%. Table 1 summarizes the proposed research papers discussed in this section.

Table 1: Literature Review Summary

Author	*Objective*	*Algorithm*	*Results*
[18]	Benchmark comparison, hybrid modeling	CNN, LSTM, RNN	BERT + Bi-LSTM 0.9046%. Bi-LSTM + CNN 0.8944%
[19]	Classifying emotions	Four-layer DNN & CNN	91% accuracy
[20]	Classifying emotions	Multi-filter CNN-Bi-LSTM	85.1% accuracy
[21]	Sentiment classification	Google BERT, attention-based Bi- LSTM, and (CNN)	BERT 64.1% BI-ATTENTIVE LSTM 60.2% CNN 59.2%
[22]	Sentiment analysis	LSTM Bi-LSTM	Bi-LSTM-based training model better than the traditional LSTM
[23]	Sentiment analysis	Bi-LSTM self-attention-based CNN	High classification accuracy and F1 measure value
[24]	Sentiment analysis	LSTM, GRU, Bi-LSTM, Bi-GRU	Bi-GRU model outperforms with an accuracy of 71.19%
[25]	Sentiment Classification	Integrating structure of CNN and Bi-LSTM	91.13% accuracy
[26]	Comparative Study Aspect-	2 interactive attention neural	BANN 73.51% 80.71%

	Level Opinion Mining	networks. Bi-LSTM CNN	CANN 69.75% 78.04%
[11]	Sentiment classification	CNN, LSTM, GRU and CNN-LSTM	CNN+LSTM 86.94 % CNN 88.72 % LSTM 86.2 % GRU 84.18 %
[27]	Sentence-level sentiment classification	Bi-LSTM-CRF	Multiple datasets 82 .3 % 48 .5%, 88 .3% 85 .4%
[28]	Document-level sentiment analysis	CNN-Bi-LSTM French language	90.66%
[29]	Aspect Based Sentiment Analysis	CNN-Bi-LSTM	3 datasets 83.21% 78.55% 73.31%.
[30]	Sentiment analysis	RNN-based LSTM and Bi-LSTM	LSTM 90.59% Bi-LSTM 90.83%
[31]	Sentiment analysis	2 BI-RNN+ VADER	BI-RNN 93%

3. Methodology

Here we will discuss the detailed approach that we followed to discover the multi-class classification of people's emotions during conflict.

3.1 Data acquisition, cleaning and preprocessing

In this study, tweets were analyzed from the Twitter website during the conflict between Ukraine and Russia. The data set consisting of 200,567 tweets were extracted from Kaggel.com. [32]. At this stage, the collected data was prepared for use in training and validating the developed model. Tweets were cleaned by applying preprocessing techniques. The following steps were taken in order to achieve this goal: In order to train and validate the developed model, the downloaded dataset had to be explored and improved. By using this method, you can find duplicates, irrelevant, and null values in your data. In accordance with the goal and problem domain, NAN values will be eliminated. Table 2 gives some examples of clean data.

Table 2: Original vs cleaned Tweets

Tweet	*Clean text*
Click on:-https://t.co/sTxbFtEIcq Clicks #androidguru9211 #Sweepstakes #win #free #follow #XRPCommunity #XRPUSD #xrpthestandard #WoodConlan #WWE2K22 #winmetawin #Wordle #Valimai #UmarRiaz #UkraineWar #Ukraineï¸• #UttarPradeshElections #USA #ONEPIECE #okullar #P2E #PAKvAUS http	click on click
Everyday the news from #Ukraine shocks the world. How can Russia allow this to happen..how can the world allow this to happen? http	everyday news shock world how russia allow happen how world allow happen

The core of a sentiment analysis system is the NLP phase. Pre-processing helps to increase the accuracy of the model through NLP and text analysis such as stemming and tokenization. In addition to these techniques, dealing with social media data requires the use of data preprocessing techniques that are more related to social media/social network data. Figure1 illustrates the steps in more detail in the following section.

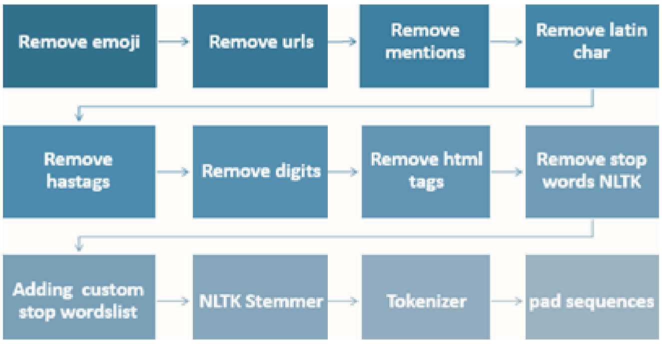

Figure 1: Preprocessing pipeline (NLP)

1) Eliminate emojis, URLs, mentions, Latin characters, hashtags, mentions, digits, and tags.
2) NLTK library used to remove stop words, add custom stop words and stemmer (by using Snowball Stemmer.) Keras used for tokenizing words and pad sequences. The tokenization method focuses on converting text into tokens until it becomes vectors. This step is essential to remove unwanted words, as shown in the next steps:

- REPLACE ALL NAN VALUES TO EMPTY STRINGS USING REPLACE() METHOD IN PYTHON
- USE NLTK LIBRARY TO LOAD STOP WORDS FOR ENGLISH
- REMOVE STOP WORDS USING THE NLTK PYTHON LIBRARY
- ADD A CUSTOM STOP WORDS LIST
- PYTHON REGEX COMPILE TO REMOVE EMOJI
- REGEX PYTHON FUNCTIONS USING SUB-FUNCTION REPLACE ONE OR MANY MATCHES WITH A STRING
- REMOVE (URLS, LATIN LETTERS, CHARACTERS, HASHTAGS, DIGITS, HTML TAGS, AND PUNCTUATIONS)
- USING NLTK LIBRARY "SNOWBALL STEMMER" STEMMER
- KERAS USED FOR TOKENIZED WORDS AND PAD SEQUENCES.

3.2 Data Visualization Tools

In this study, data were analyzed in various ways to determine how they were related. Our visualizations used several Python pre-built libraries. Using bar graphs, line graphs, and word clouds, we visualized the data.

3.3 Classification

In our research, we proposed a simple and efficient Deep learning-based classification model using a single

Bidirectional LSTM network for English sentiment analysis. A multi-class classification approach (positive, negative and neutral sentiment of tweets) to classify people's emotions during the war showed that English sentiment analysis performed best when a single Bidirectional LSTM was used. The architecture of our proposed model is shown in Figure 2. It is comprised of an embedding layer where every index, corresponding to a unique word in the data set, is transformed into a real-valued feature vector. The Bi-LSTM layer is capable of reading input reviews in both directions, forward and backward. The output of Bi-LSTM can be summarized by concatenating the forward and backward states, then performing pooling to reduce the number of parameters and computations. Dense Layer will describe how the neurons are connected to the next layer of neurons which works for changing the dimension of the output by performing matrix vector multiplication. Finally, the Dropout layer will remove the noise that may be present in the input of neurons. This prevents overfitting the model.

Figure 2: The architecture of the proposed Model

4. Experimentation and Results

This section contains information system configuration, data analysis, and performance metrics, results, and comparative analysis.

4.1 Train the Model

Training and testing data sets were split 80:20 for this study. In training, we make a classification based on how the network is performing. We calculate how incorrect the classification is. This error is then minimized by updating the weights or parameters of the network. Until the model stops adapting, the procedure is repeated. This procedure requires three important parameters to be chosen.

1) Metric: Accuracy report and confusion matrix used in the experiment to measure model performance.
2) Loss function: Calculates a loss value, which is then minimized by tuning the weights of the network during the training phase. Multiclass classification models use categorical_crossentropy as a loss function.
3) Optimizer: Function that updates network weights based on loss function output. Our experiments were conducted using the Adam optimizer.

4.2 Tune Hyper parameters

Neural networks are capable of learning complicated relationships between their inputs and outputs. There is a possibility that many of these relationships are caused by sampling noise. Overfitting may result from this problem. And thereby decreases the model's capacity for classification. Overfitting was reduced by using dropout layers. Adapting these hyper parameters to the particulars of each problem will help refine our model for a better representation of it. For further explanation, we discuss below the most appropriate hyper parameters for our proposed model. Table 3 lists the values of the hyper parameters.

The number of layers determines the complexity of the network. This value must be carefully considered. An over fitted model will be able to learn too much information about the training data by using too many layers. If the model has too few layers, under fitting can occur. For text classification datasets, we experimented with a single BI-LSTM layer.

Table 3: Optimal hyper parameters of the proposed model

Hyper parameters	*Value*
Optimizer	Adam
Loss function	categorical_crossentropy
Batch size	64
Epochs	4
Dropout rate	0.25
Bi-LSTM Nodes	64
Max length	300
Learning rate	0.001
Regularizes	L2

- Number of units per layer: Each layer must contain the information it needs to perform its transformation. 32/64 units performed well.
- Dropout rate: Dropout layers are used in the model for regularization. They define the fraction of input to drop as a precaution against overfitting. We applied 0.25.
- Embedding dimensions: For word embedding, the size of each vector determines the number of dimensions. An embedding dimension of 100 was used.

5. Data Analysis

5.1 Tweets Sentiment Analysis

Sentiment analysis allows researchers to find out about the aims, messages, and impacts of social media content. Visual representations of sentiment analysis are provided by word clouds. We used a word cloud to classify tweets into three categories, in order to determine whether they were influenced by psychological factors.

5.2 Sentiment words according to polarities

The numbers of tweets classified as positive, negative, and neutral are shown in Figure 3. It is observed that 15.27% of tweets had positive sentiments, 36.96% were having neutral sentiment and 47.77% tweets were having negative sentiments. As compared to negative and neutral sentiments, more people expressed negative sentiment. Negative tweets formed the majority. It was less common for people to react positively. The majority of people remained negative about the situation. It is evident from negative tweets that the situation has taken a serious turn in the state and the emotions of the people.

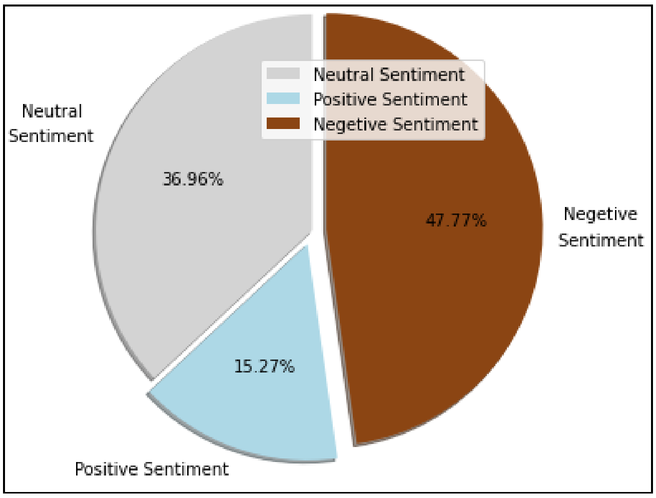

Figure 3: Tweets Sentiment percentage

5.3 Word Cloud

A word cloud visualizes the most common words in a dataset. The most frequently used words are represented by the entire word cloud. More often than words with smaller fonts, words with larger fonts are used. The word cloud can provide us with an overview of people's emotions during conflict. Figure 4 depicts positive sentiments. Figure 5 shows the negative sentiment words, and the neutral sentiment words are shown in Figure 6.

Figure 4: Frequent words visualization for positive words

Figure 5: Frequent words visualization for negative words

Figure 6: Frequent words visualization for neutral words

5.4 Most Frequently Occurring Words

Figure 7 shows Most Frequently Occurring Words - Top 15 which are Putin, Ukrain, Support, Presid, Today, Kyiv, need, doh, help, forc, minsit, world, stand, us and invas.

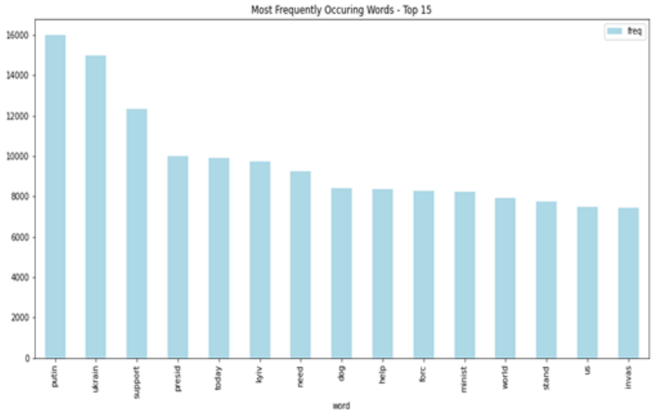

Figure 7: Most Frequently Occurring Words - Top 15

5.5 Hashtags

Hashtags count per tweet. Most Twitter users utilize hashtags extensively, and they are sometimes associated with important events or trends. The total number of hashtags is plotted on the x-axis, and the most frequently used hashtags are plotted on the y-axis, illustrating the density of tweets. The results are presented in Figures 8 ,9 ,10 respectively.

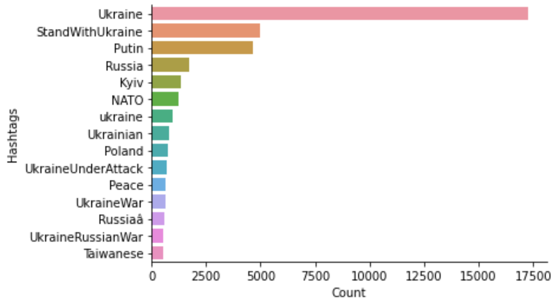

Figure 8: Positive Hashtags

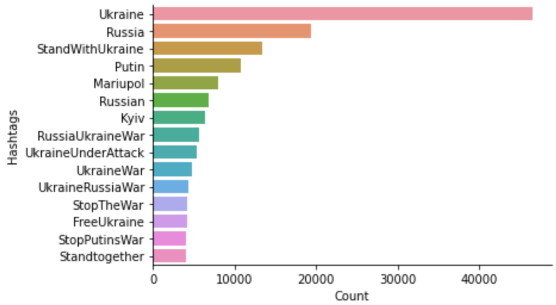

Figure 9: Negative Hashtags

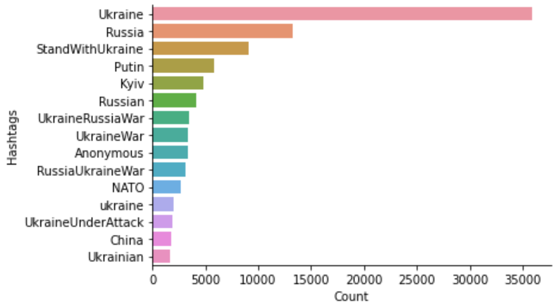

Figure 10: Neutral Hashtags

6. Evaluation

A performance metric includes accuracy, precision, recall, and F1-Score. The next subsection briefly explains the state-of-the-art evaluation techniques.

6.1 Accuracy Report

The effectiveness of our model is evaluated by accuracy report, the precision, recall and F1 score were calculated. The precision represents the prediction percentage, the recall represents the percentage of positive cases caught, F1-score represents the correctness percentage of positive predictions, and support indicates how many times a particular class actually occurs in the specified dataset. Table 4 illustrates the results. The performance of the proposed architecture layers has been compared with the state-of-the-art models LSTM+CBA+LA, CNN+LSTM, WALE+LSTM, FARNN-ATT and single BI-LSTM in terms of accuracy [33]. Table 5 will illustrate how the proposed model performed compared with other models in terms of accuracy. It is evident that our model performs better than others English data [33] in term of Accuracy with 91.79%.

Table 4: Accuracy report for each class

Class	*Precision*	*Recall*	*F1-Score*	*Support*
Neutral	0.91	0.88	0.90	14,617
Positive	0.89	0.89	0.89	6,097
Negative	0.94	0.94	0.94	19,398

Table 5: Result comparisons of English tweets dataset with other methods

Method	*Accuracy%*
LSTM+CBA+LA	90.10
CNN+LSTM	88.90
WALE+LSTM	89.50
FARNN-ATT	89.22
Single BI-LSTM	90.585
Proposed Model	91.79

6.2 Training and validation accuracy

Epochs were set to 20, but an early stop callback with patience 4 was used on validation loss save time and prevent over-fitting. Epochs are the number of passes of the entire data during training. Patience is the number of epochs when there is no further improvement and hence the training process finishes. Model quality is determined by the corresponding errors or losses. For example, a high validation loss would signify that the trained model does not generalize well to the validation data. This means it is not good enough for the testing phase. The results for training and validation accuracy and model loss accuracy are presented in Figures 11, 12 respectively.

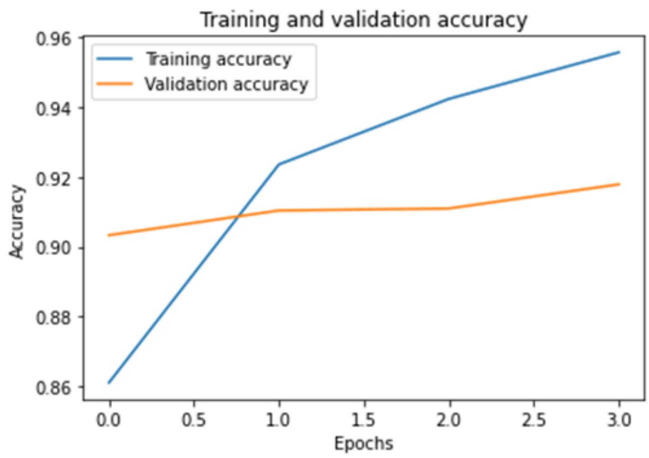

Figure 11: Training and validation accuracy

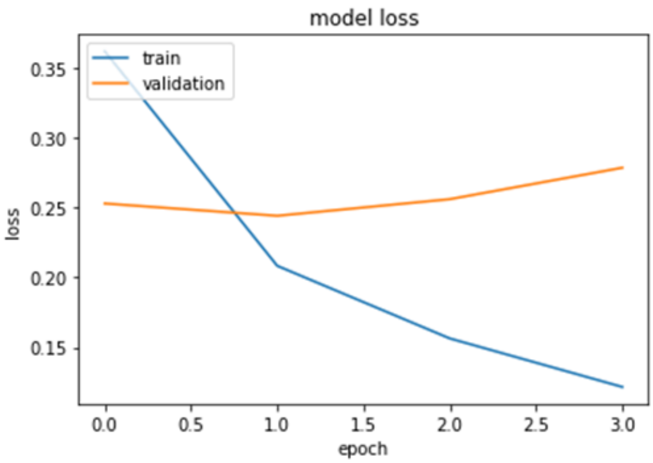

Figure 12: Model Loss

6.3 Confusion Matrix

A classification's accuracy is assessed using a confusion matrix. Predicted labels are on the x-axis, while the true labels of our test set's samples are on the y-axis. A prediction was made based on the test set. Figure 13 depicts the confusion matrix. The true label represents the tweet's actual sentiment, while the predicted label represents its predicted sentiment. The confusion matrix implies that 13, 032+5, 478+18, 317 = 36,827 tweets were predicted with the same sentiment as their true label. It also implies that 555+1040+ 496+ 123+919+162 = 3,295 tweets were predicted with the wrong sentiment. As predicted by their true label, 91.8% of the tweets had a correct sentiment, while the rest had the wrong sentiment.

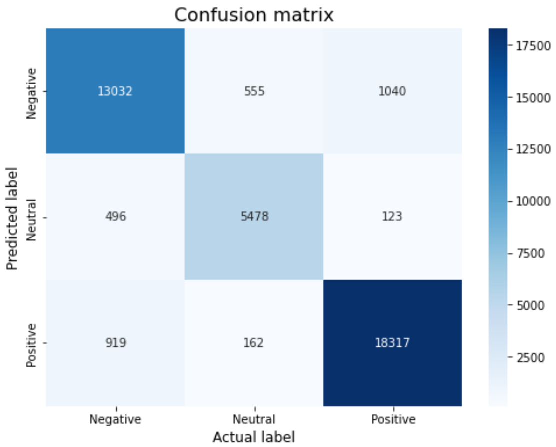

Figure 13: Confusion Matrix

7. Discussion

In this study, we presented a simple and efficient model using a single Bidirectional LSTM network. From the analysis report it is evident that conflict between Ukraine and Russia has hugely impacted the lives of people, and people are badly affected by the war, but most people are still positive in handling the war. People with Negative tweets can be marked as vulnerable and report of the same should be transmitted to the appropriate authorities. Our approach achieved competitive results. The results are substantially better than several novel methods in terms of accuracy, the results mentioned in this study demonstrated that it can be possible to use much simpler architecture to achieve the same level of classification performance.

An overall accuracy of 91.79% shows that the classifier did well in classification the sentiment polarities. Model showed good prediction scores in precision, recall, F-1 scores, and confusion matrix calculation. The confusion matrix offers a detailed of how the classifier did the classification with respect to the positives, negatives and neutral. that was used to substantiate the various percentages of tweets that positive, negative sentiments and neutral., Moreover it can be inferred that 47.77% of the tweets were negative with 36.96% and 15.77 % for the neutral.

8. Conclusion and Future work

To design an efficient and mature sentiment analysis method, many obstacles must be addressed. The majority of these challenges are inherited from datasets like the small dataset size which is used in this study, a larger dataset has more probability of giving better results. This is because the model will get more opportunities to learn, which could lead to even better accuracy. Few main contributions that help to discover people attitudes or sentiments in various cases when war happening around the world. The study presented a detailed analysis of sentiments expressed by people during war and enabled us to understand the emotions and needs of people during conflict.

Understanding emotions will help to overcome and prevent misunderstandings, loss of opportunities, and sometimes violence. Negative emotions, such as anxiety, which arise during a conflict and lead to other mental health outcomes, are greatly reduced when negative emotions are eliminated [3]. Elimination needs careful understanding and analysis.

Our results confirm that the percentage of negative emotions is 47.77%. The paper suggests improving the attention of people during war to their mental health. It is crucial to develop technologies that support people during conflicts like online clinics. This is because it is very difficult to move around during conflicts or other crises. By drawing attention to emotions during crises, it will gives insights to maintain mental health and wellness. This will help keep our lives healthy and safe. The framework we present provides robust explanation and classification capabilities. This paper clearly illustrates that it is possible to use a much simpler architecture to achieve the same level of classification performance. The proposed achieved competitive results and well performed several novel methods in terms of accuracy.

This study will consider multi-class sentiment classification as well as a multilingual approach with more languages in order to evaluate the shifting emotions and feelings of individuals over time, and also to ascertain if there are any noticeable changes in their mental health over time. Therefore, we intend to extend this study to Arabic text data to provide a more comprehensive explanation of sentiment classification.

References

[1] “Understanding Your Emotions (for Teens) - Nemours KidsHealth.” https://kidshealth.org/en/teens/understand-emotions.html (accessed Nov. 30, 2022).
[2] D. Grewal and P. Salovey, “Feeling Smart: The Science of Emotional Intelligence,” Am. Sci. - AMER SCI, vol. 93, Jul. 2005, doi: 10.1511/2005.54.969.
[3] M. Pettinelli, The Psychology Of Emotions, Feelings and Thoughts. Mark Pettinelli, 2007.
[4] B. Liu, Sentiment Analysis and Opinion Mining, vol. 5. 2012. doi: 10.2200/S00416ED1V01Y201204HLT016.
[5] C. Zucco, B. Calabrese, and M. Cannataro, Sentiment analysis and affective computing for depression monitoring. 2017, p. 1995. doi: 10.1109/BIBM.2017.8217966.
[6] E. H. Hovy, “What are Sentiment, Affect, and Emotion? Applying the Methodology of Michael Zock to Sentiment Analysis,” in Language Production, Cognition, and the Lexicon, N. Gala, R. Rapp, and G. Bel-Enguix, Eds. Cham: Springer International Publishing, 2015, pp. 13–24. doi: 10.1007/978-3-319-08043-7_2.
[7] S. Symeonidis, D. Effrosynidis, and A. Arampatzis, “A comparative evaluation of pre-processing techniques and their interactions for twitter sentiment analysis,” Expert Syst. Appl., vol. 110, pp. 298–310, Nov. 2018, doi: 10.1016/j.eswa.2018.06.022.
[8] A. Kadhim, “An Evaluation of Preprocessing Techniques for Text Classification,” Int. J. Comput. Sci. Inf. Secur., vol. 16, pp. 22–32, Jun. 2018.
[9] J. Camacho-Collados and M. T. Pilevar, “On the Role of Text Preprocessing in Neural Network Architectures:

An Evaluation Study on Text Categorization and Sentiment Analysis," Jul. 2017.

[10] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. 2018.

[11] O. Hourrane, N. Idrissi, and E. H. Benlahmar, An Empirical Study of Deep Neural Networks Models for Sentiment Classification on Movie Reviews. 2019, p. 6. doi: 10.1109/ICSSD47982.2019.9003171.

[12] S. Sukheja, S. Chopra, and M. Vijayalakshmi, "Sentiment Analysis using Deep Learning - A survey," in 2020 International Conference on Computer Science, Engineering and Applications (ICCSEA), Mar. 2020, pp. 1–4. doi: 10.1109/ICCSEA49143.2020.9132863.

[13] S. Hochreiter and J. Schmidhuber, "Long Short-term Memory," Neural Comput., vol. 9, pp. 1735–80, Dec. 1997, doi: 10.1162/neco.1997.9.8.1735.

[14] Y. Indulkar and A. Patil, Sentiment Analysis of Uber & Ola using Deep Learning. 2020, p. 27. doi: 10.1109/ICOSEC49089.2020.9215429.

[15] K. Pasupa and T. Ayutthaya, "Thai Sentiment Analysis with Deep Learning Techniques: A Comparative Study based on Word Embedding, POS-tag, and Sentic Features," Sustain. Cities Soc., vol. 50, May 2019, doi: 10.1016/j.scs.2019.101615.

[16] "The Most Spoken Languages in the World | Berlitz." https://www.berlitz.com/blog/most-spoken-languages-world (accessed Nov. 30, 2022).

[17] H. Hahn, "Correlates of Public Sentiments About War: Local Referenda on the Vietnam Issue," Am. Polit. Sci. Rev., vol. 64, no. 4, pp. 1186–1198, 1970, doi: 10.2307/1958365.

[18] C. Colón-Ruiz and I. Segura-Bedmar, "Comparing deep learning architectures for sentiment analysis on drug reviews," J. Biomed. Inform., vol. 110, p. 103539, Oct. 2020, doi: 10.1016/j.jbi.2020.103539.

[19] K. Hasib, Md. A. Habib, N. A. Towhid, and Md. I. H. Showrov, A Novel Deep Learning based Sentiment Analysis of Twitter Data for US Airline Service. 2021. doi: 10.1109/ICICT4SD50815.2021.9396879.

[20] Z. Rajabi, A. Shehu, and O. Uzuner, Beyond Binary Sentiments: A Multi-Channel BiLSTM-CNN model for Multilabel Emotion Classification of Informal Text. 2020. doi: 10.1109/ICSC.2020.00060.

[21] A. Roy and M. Ojha, Twitter sentiment analysis using deep learning models. 2020, p. 6. doi: 10.1109/INDICON49873.2020.9342279.

[22] L.-C. Chen, C.-M. Lee, and M.-Y. Chen, "Exploration of social media for sentiment analysis using deep learning," Soft Comput., vol. 24, Jun. 2020, doi: 10.1007/s00500-019-04402-8.

[23] P. Bhuvaneshwari, A. Rao, Y. Robinson, and M. Thippeswamy, "Sentiment analysis for user reviews using Bi-LSTM self-attention based CNN model," Multimed. Tools Appl., vol. 81, Apr. 2022, doi: 10.1007/s11042-022-12410-4.

[24] S. Sachin, A. Tripathi, N. Mahajan, S. Aggarwal, and P. Nagrath, "Sentiment Analysis Using Gated Recurrent Neural Networks," SN Comput. Sci., vol. 1, Mar. 2020, doi: 10.1007/s42979-020-0076-y.

[25] S. Tam, R. Ben Said, and O. Tanriover, "A ConvBiLSTM Deep Learning Model Based Approach for Twitter Sentiment Classification," IEEE Access, vol. PP, pp. 1–1, Mar. 2021, doi: 10.1109/ACCESS.2021.3064830.

[26] W. Quan, Z. Chen, J. Gao, and T. Hu, Comparative Study of CNN and LSTM based Attention Neural Networks for Aspect-Level Opinion Mining. 2018, p. 2150. doi: 10.1109/BigData.2018.8622150.

[27] T. Chen, R. Xu, and X. Wang, "Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN," Expert Syst. Appl., Nov. 2016, doi: 10.1016/j.eswa.2016.10.065.

[28] M. Rhanoui, M. Mikram, S. Yousfi, and S. Barzali, "A CNN-BiLSTM Model for Document-Level Sentiment Analysis," Mach. Learn. Knowl. Extr., vol. 1, pp. 832–847, Jul. 2019, doi: 10.3390/make1030048.

[29] W. Meng, Y. Wei, P. Liu, Z. Zhenfang, and H. Yin, "Aspect based Sentiment Analysis with Feature Enhanced Attention CNN-BiLSTM," IEEE Access, vol. PP, pp. 1–1, Nov. 2019, doi: 10.1109/ACCESS.2019.2952888.

[30] K. Nabiul Alam et al., Deep Learning-Based Sentiment Analysis of COVID-19 Vaccination Responses from Twitter Data. 2022. doi: 10.48550/arXiv.2209.12604.

[31] I. López Ramírez and J. Méndez Vargas, A sentiment analysis of the Ukraine-Russia conflict tweets using Recurrent Neural Networks. 2022.

[32] "EDA_and_Sentiment_Analysis." https://kaggle.com/code/ssaisuryateja/eda-and-sentiment-analysis (accessed Nov. 30, 2022).

[33] Z. Hameed and B. Garcia-Zapirain, "Sentiment Classification Using a Single-Layered BiLSTM Model," IEEE Access, vol. 8, pp. 73992–74001, 2020.