Expert Systems With Applications 284 (2025) 127759

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Words of War: A hybrid BERT-CNN approach for topic-wise sentiment analysis on The Russia-Ukraine War

Md. Saiful Islam, Mahmuda Ferdusi, Tanjim Taharat Aurpa *

Bangabandhu Sheikh Mujibur Rahman Digital University, Gazipur, Bangladesh

ARTICLE INFO

Keywords:
Russian-Ukraine war
BERT-CNN
Topic-wise sentiment analysis
VADER
Topic modeling

ABSTRACT

The Russia-Ukraine War has dramatically impacted the world, affecting economies, lives, and politics. The war is a common topic on social media, especially on platforms like YouTube. In this study, we analyzed YouTube comments from videos posted by popular news channels like CNN, BBC, etc., to understand people's opinions about the war. We used a tool called VADER for sentiment analysis and an unsupervised BERT model to identify ten key topics related to the war, including humanitarian issues, economic challenges, political debates, and societal concerns. We then created a model that combines BERT's ability to understand context with CNN's feature extraction strengths. Unlike existing approaches, our model incorporates an extra input layer that considers the topic as a significant feature. This hybrid model effectively classifies sentiments with 92.26% accuracy. Our research provides insights into public perceptions and discussions about the Russia-Ukraine War, highlighting essential themes in the conversation.

1. Introduction

The Russian invasion of Ukraine in early 2022 has triggered massive global disruptions by affecting food security, geopolitics and economic stability (Pereira, Bašić, Bogunovic, & Barcelo, 2022). The conflict has tightened the supply chains, increased commodity prices and resulted in global crisis. The financial sanctions on Russia had impacted more. It had weakened international markets. These have furthered inflation wherein many important commodities like fuel and food have seen tremendous surges in prices (Assaf, Gupta, & Kumar, 2023; Orhan, 2022). In the energy sector, crude oil prices saw sharp increases, with West Texas Intermediate (WTI) crude rising by 52.33% and Brent crude by 56.33% between October 2021 and August 2022 (Zhang, Hu, Jiao, & Wang, 2024). Similarly, maize and wheat prices rose by 27% and 13%, respectively, in January 2023 compared to January 2021 (Pietrzak, 2024).

Beside the economic consequences, the war has redrawn global diplomatic alignments and widespread public discourse. In this paper we analyzes the public sentiment on these themes by making use of data collected from YouTube comments which reflect the gap between the objective tone of news reporting and subjective public reactions. This kind of social media data captures diverse perspectives, thereby enriching policy discussions and fostering better understanding of global sentiment about the Russia-Ukraine war crisis.

Traditional methods of sentiment analysis often fail to capture the complexity of public opinions about multi-faced issues like war. Sentiment analysis in this study shown more nuanced exploration of public opinion, examining elements such as sentiment polarity, opinion terms and prevailing attitudes toward topics like military actions and humanitarian concerns (Zhang, Li, Deng, Bing, & Lam, 2022).

Given the vast amounts of unlabeled social media data, unsupervised sentiment analysis is essential for large-scale studies. Unsupervised clustering techniques, particularly suited to identifying hidden themes/trends within massive, unlabeled datasets, enable this research to classify discussion topics without manual labeling is a critical factor when analyzing vast social media datasets (Hu, Tang, Gao, & Liu, 2013). In this study, we use unsupervised BERT based topic modeling to identify main themes in the discourse, complemented by sentiment analysis using the VADER lexicon-based model, which is effective for social media language and BERT for fine-grained sentiment detection (Pano & Kashef, 2020; Wada et al., 2024).

More recently the BERT is a pre-trained transformer model with self-attention mechanisms that has gained good recognition in natural language processing tasks for its accuracy regarding text based categorization and classification. The use of BERT based classification with VADER allows this study to find the hidden public sentiment with greater precision and accuracy (Aurpa & Ahmed, 2024).

These advanced methods of sentiment analysis aim at providing an overall view of global public opinion on the Russia-Ukraine war, contributing value to possibly shape future discussions and policy decisions.

* Corresponding author.
E-mail addresses: 2002047@icte.bdu.ac.bd (Md. Saiful Islam), 2002010@icte.bdu.ac.bd (M. Ferdusi), aurpa0001@bdu.ac.bd (T.T. Aurpa).

https://doi.org/10.1016/j.eswa.2025.127759
Received 25 January 2025; Received in revised form 29 March 2025; Accepted 13 April 2025
Available online 29 April 2025

1.1. Research objectives

The paper tends to unfold the emotions within the YouTube comments on the Russia-Ukraine War, showing the major trends and themes lying underneath them. The analysis will range from locating immediate emotional responses to how these sentiments are evolved over time in relation to significant events in the conflict. In this research, we are analyzing the words of public from their responses on YouTube video. These words related to this war are used to understand underlying topics and people's sentiments. Thus, we have named this work "Words of War". This study will apply advanced sentiment analysis techniques, especially VADER and sentiment analysis, to assess the subtlety of public opinion and its implications for the broader discourse on the war. The specific objectives of this research are summarized below.

- Analyzing Different social media comments to observe public opinion on this topic.
- Proposing a hybrid methodology using BERT and CNN architectures that utilizes all significant features of the dataset.
- Analyzing the Public's sentiment on different sub-topics related to the Russia-Ukraine War.
- Evaluating the proposed model with different Evaluation Metrics.

2. Related work

Sentiment analysis merges data and text mining methodologies to uncover sentiments embedded within written text, serving as a foundation for a range of emotion and opinion studies (Wadhwani, Varshney, Gupta, & Kumar, 2023).

In the case of searching for emotion in English tweets using a single bidirectional LSTM network on the Russia-Ukraine War, It is proposed by a paper that states working with multi-class classification can work for such purposes positive, negative and neutral being the classes here. By combining Bidirectional Transfer Long Short Term Memory with global Max pooling ID Mechanism, we achieved an accuracy of 91.79% using only one Bi-LSTM Layer (Shlkamy, Mahar, & Sedky, 2023).

Previous research has explored both sentiment and emotion detection but improvements remain, specially in identifying sentiments related to racism and the Russia-Ukraine conflict (Al Maruf, Ziyad, Haque, & Khanam, 2022). For instance, an analysis of tweets tagged #UkraineRussia on the first day of the conflict revealed frequently occurring words like "war", "people", "world", "Putin" and "peace" (Hasan, Islam, Jahan, Meem, & Rahman, 2023). Another study applied the Word Cloud technique, highlighting terms such as "Russia", "Russian", "Ukraine", and "Ukrainian" to visualize commonly used words (Aslan, 2023). A recent sentiment analysis of 2537 Indonesian tweets about the conflict demonstrated an accuracy of 82%, classifying 54.7% as positive, 35% neutral and 10.2% negative sentiments (Simarmata et al., 2023).

Sentiment analysis uses a lexical function named VADER which is particularly effective at sentiments posted on social media text over the Russia-Ukraine War (Vaghela, Makwana, Chande, & Mehta, 2024). Flair sentiment labels showed a more negative tone compared to the others. TexBlob resulted in more neutral and positive feelings. VADER sentiment produced a negative mood (Scheerder). In this paper the author used Valence Aware Dictionary and Sentiment Reasoner (VADER) to prepare the posts we have gathered and figure out if they are neutral, positive or negative. For getting know the impact of the Russia-Ukraine War among the netigens (Krivičić & Martinčić-Ipšić, 2023). Another analysis on the Russia-Ukraine war the results showed that most English tweets convey fear and anger as predominant feelings, reaching 32.08% and 15.18% of the total tweets analyzed, respectively. Regarding tweets in Russian, the majority presented negative polarity, with 86.83% of the total comments (Ramos & Chang, 2023).

Research on Twitter data has also utilized hybrid approaches, such as combining VADER, GloVe embeddings and deep learning techniques, achieving a high accuracy of 97.09% in labeling 1 million tweets collected from January to February 2023 (Sinha, Innani, Chinnari, & Khan, 2023). Sentiment expressed in Russian-language posts was also analyzed across VK and Telegram, with 1,393,245 conflict-related posts showing sentiment trends that evolved over time from February 2022 to September 2022 (Dean & Porter, 2024).

BERT (Bidirectional Encoder Representations from Transformers) is a transformer-based model developed by Google for natural language processing tasks. After 10 iterations, the BERT model achieved an accuracy of 86.43% of twitter based analysis of the Russia-Ukraine War (Maheshwari, Chandra, Yadav, & Gupta, 2023).

In recent times, The hybrid models have gained significant attention in the field of Machine Learning and Natural Language Processing due to their ability to combine different approaches which helps to improve accuracy and interpretability (Bauskar, Badole, Jain, & Chawla, 2019; Salur & Aydin, 2020). The Siagra-ConSA-HSOA method, combining sentiment analysis from Twitter with stock market data and a Hybrid Swarm Optimization Algorithm (HSOA), achieves outstanding results with 99.89% accuracy (Kumarappan, Rajasekar, Vairavasundaram, Kotecha, & Kulkarni, 2024). Another Hybrid Model: BERT with rule-based method has been proposed which gained 76% accuracy detecting emotions (Madhuri & Lakshmi, 2021). Hybrid model using generative autoencoders and ensemble learning was developed to improve lung image diagnosis, enhancing accuracy with attention mechanisms (Rajasekar, Chandra, Pears, Vairavasundaram, & Kotecha, 2025). We intend to apply a hybrid approach in this research to observe the remarkable performance of different hybrid approaches and analyze the corpus.

Table 1
Dataset Summary.

Attribute	Description
Total Comments	85,904
Time Period	January 2022 – August 2024
News Channels	BBC, CNN, Al Jazeera etc

3. Methodology

3.1. Data collection

The dataset used for this study is the comments. gathered from different YouTube news channels such as BBC, CNN, Al Jazeera on the Russia-Ukraine War, within the time span of 2022 to Aug 2024. In all, a total of 85904 comments were gathered, which is representative of the diverse range of opinions and sentiments regarding the Russia-Ukraine war. Table 1 provides an overview of the dataset.

To collect the comments, we first extracted data from videos uploaded on YouTube. We collected comments from YouTube for getting News Centric Discussions and relevant comments. Less noise BOTS and Spams comparing to other social media. Audio-Visual influence on public sentiments by NEWS Media Reports. However, the extracted comments often contain irrelevant noise and symbols that could affect the subsequent analysis. Therefore, a pre-processing phase is necessary to clean the dataset and enhance the quality of the data used for sentiment analysis (Pradhan, 2021). Here data collection and Preprocessing explained in Fig. 1.

To further validate our methodology, we applied it to different social media datasets collected from Kaggle. Specifically, we analyzed a Twitter dataset (see 3.1) and a Reddit dataset (see 3.1), both of which focus on the Russia-Ukraine conflict.

Twitter Dataset: The Twitter dataset was collected from Kaggle and contains tweets related to the Russia-Ukraine conflict. The dataset was consisting 17951 tweets. The dataset can be accessed at:[1]

[1] https://www.kaggle.com/datasets/vanamayaswanth/russia-vs-ukraine-tweets

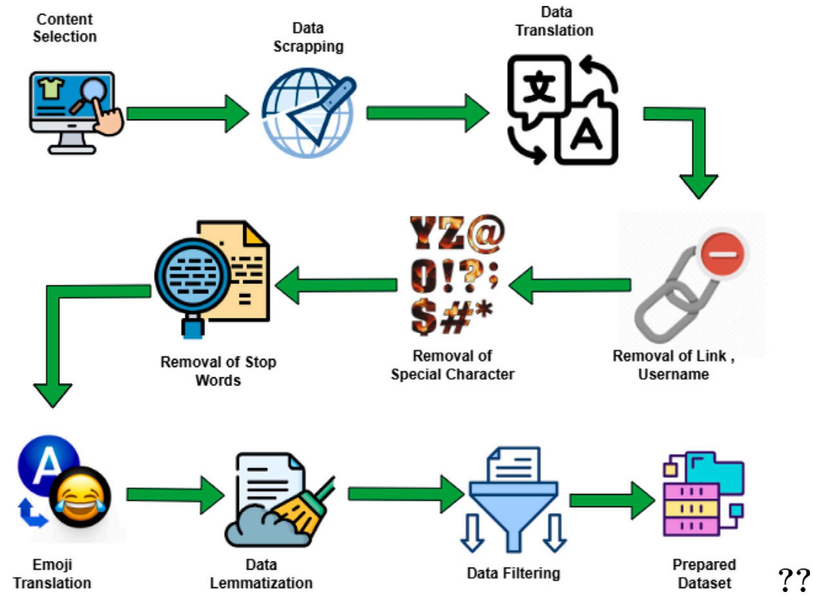

Fig. 1. Data Collection and Preprocess.

Reddit Dataset: The Reddit dataset was also obtained from Kaggle and includes comments from Reddit discussions on the conflict. The dataset includes was 64906. The dataset can be found at:[2]

For both datasets, we applied the same preprocessing and analytical steps as in our primary dataset to ensure consistency. This comparison will allow us to evaluate the effectiveness and generalizability of our approach across different social media platforms.

3.2. Data preprocess

The data preprocessing phase is very important to assure the quality and reliability of sentiment analysis. In this work, a set of manual checks were conducted over the collected YouTube comments to remove irrelevant contents and enhance the overall integrity of the dataset (Vaghela et al., 2024).

Preprocessing is indispensable because raw data often contains noise, inconsistencies and irrelevant information that can negatively impact the performance of machine learning models and the overall evaluation process. YouTube comments scraped contain a lot of spam, ads and irrelevant discussions that carry absolutely no value for sentiment analysis results. Grammatical errors, typos, special characters and URLs are bound to be present in text data, which needs to be cleaned for meaningful processing (Ahmed, Aurpa, & Anwar, 2020). Lack of preprocessing will result in biases being introduced, increased sparsity in data and deterioration of model performance (Uysal & Gunal, 2014). This would ensure that the dataset is accurate and valid by preprocessing (Haddi, Liu, & Shi, 2013), hence making the results of further analysis robust and reliable.

The preprocessing steps included the following:

1. **Language Translation :** After collecting comments there were many comments were in different language. We have translated the comments into English for applying our models on dataset.

2. **Removal of Links, email and username :** Any hyper- links present in the comments were removed to avoid the analysis being biased by irrelevant web references. As well as email and username also removed from the dataset.

3. **Removal of Special Character :** We also removed the special character occurred in the comments, which are irrelevant to our work (Pradha, Halgamuge, & Vinh, 2019).

4. **Stopword Removal :** Common stopwords, which do not contribute meaningful information to Sentiment Analysis, were removed from the comments. This step helps in removing noise and retaining only the substantial words that convey meaning (Vijayarani et al., 2015).

5. **Emoji Translation :** Emojis can convey sentiments which are often missed when using plain text. Thus, a translation approach to convert emojis into their textual equivalent is adopted. A translation process was applied in order to do justice regarding the sentiment the symbol conveys in the analysis (Liu et al., 2021).

6. **Lemmatization :** This constitutes the most crucial textual preprocessing. Lemmatization is the technique NLP uses to reduce words to their base or root form-the lemma. Unlike stemming, which usually simply truncates words based on simple rules, lemmatization takes into account the context and morphological analysis of words with the aim of ensuring that a valid word in the language is obtained from the derived lemma (Chauhan et al., 2023).

Each of the steps in preprocessing was important in preparing the dataset to conduct effective sentiment analysis, which assured that results are representative of the actual public opinion about the Russia-Ukraine war.

These preprocessing steps were essential to prepare the dataset for an effective sentiment analysis and to ensure that the results reflect genuine public sentiment regarding the Russia-Ukraine conflict.

3.3. Data labeling

We have used two different types of data labeling here. The data labeling techniques are given below:

3.3.1. Topic modeling

The unsupervised clustering was done as part of our analysis, where a BERT-based model was used to perform unsupervised clustering for topic modeling, identifying ten relevant topics in the dataset. At the same time, BERT allows us to find comments that semantically may group into themes nicely, thus increasing interpretability. The following topics are identified in the dataset: Geopolitics, Civilian, Putin, War Strategy, Children of War, USA Politics and NATO, News and Media, Economy, World War 2- World War 3 (WW2 and WW3) and War

[2] https://www.kaggle.com/datasets/tariqsays/reddit-russiaukraine-conflict-dataset

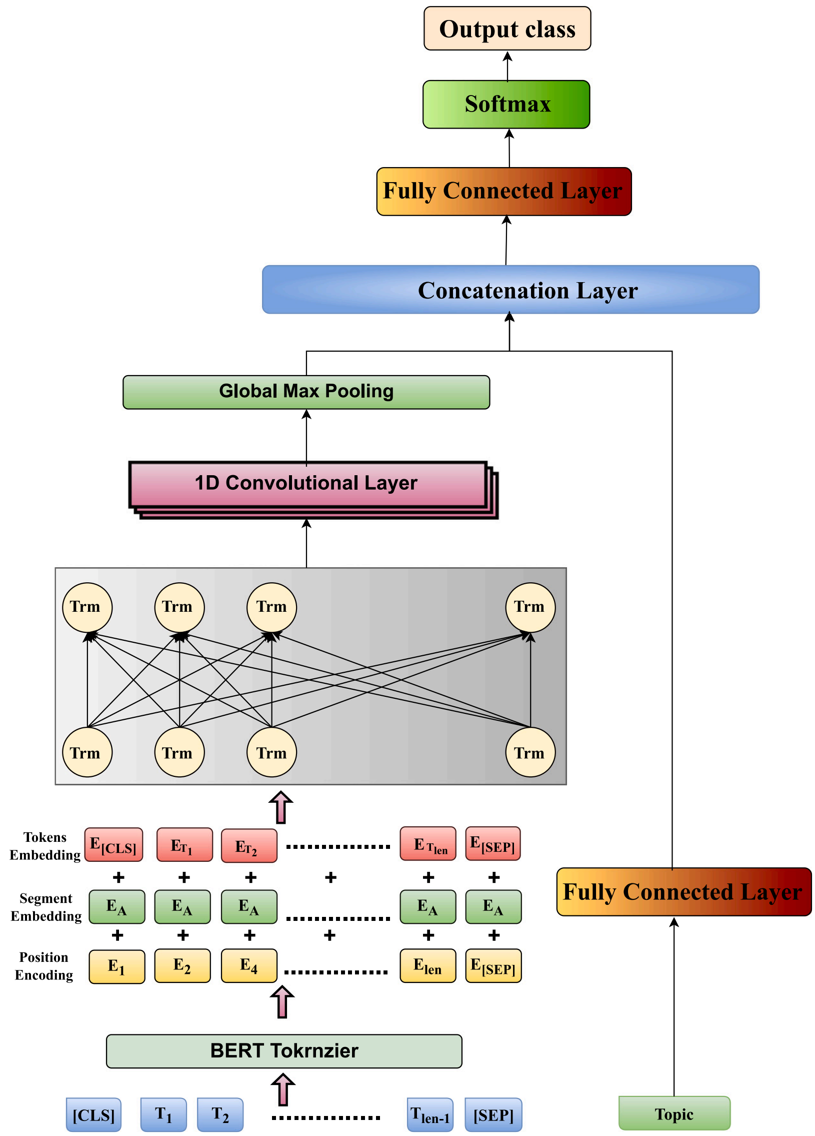

Fig. 2. Proposed BERT-CNN Model for topic wise Sentiment Classification.

Crimes. Each of these topics is representative of a domain of discourse that is particularly relevant in the global situation brought about the Russia-Ukraine war, hence bringing in different aspects ranging from political strategies to humanitarian impact and further economic consequences. This topic classification is basic in assessing the thematic structure of comments and serves as a very valuable contribution to structured sentiment analysis.

3.3.2. Sentiment label identification

For the sentiment analysis, the comments were rated using VADER, which is a lexicon and rule-based sentiment analysis (Jeba, Aurpa, Siyam, Khan, & Mansia, 2023). It is totally adapted for the analysis of social media content and, therefore, it was particularly effective in this dataset. Sentiment coupled with a score was rated for each comment, which was labeled as either Positive, Negative or Neutral. VADER's ability to process emotive language and informal expressions enriches the capture of nuances in user sentiment, so crucial in analyzing social media discourse. This annotation provides a quantitative measure of emotional responses for each topic, therefore allowing the comprehensive analysis of public opinion and dynamic sentiment in relation to the Russia-Ukraine war.

3.3.3. Humanize evaluation

Simple Random Sampling is one of recognized process for the evaluation. It is considered to be very effective where each unit is uniformly or closed to be selected (Noor, Tajik, & Golzar, 2022).

To ensure accuracy, we manually evaluated a sample of 12,000 randomly selected comments from the dataset, with 2000 comments. The sentiment labels (positive, neutral and negative) were cross-checked to confirm their relevance to the comment content and we found the comments and sentiments are relevant. The reason we have selected Simple Random Sampling is to ensure every comment had a equal chance to be selected. This approach minimizes bias and provides a representative sample of the complete dataset.

3.4. Model selection

3.4.1. BERT

The BERT model is a type of pre-trained language model introduced by Google, which changed NLP due to its effective potential to understand context from either direction. This differs from earlier models that processed language in sequence. BERT processes language in both directions at all levels and captures richer context with more nuanced

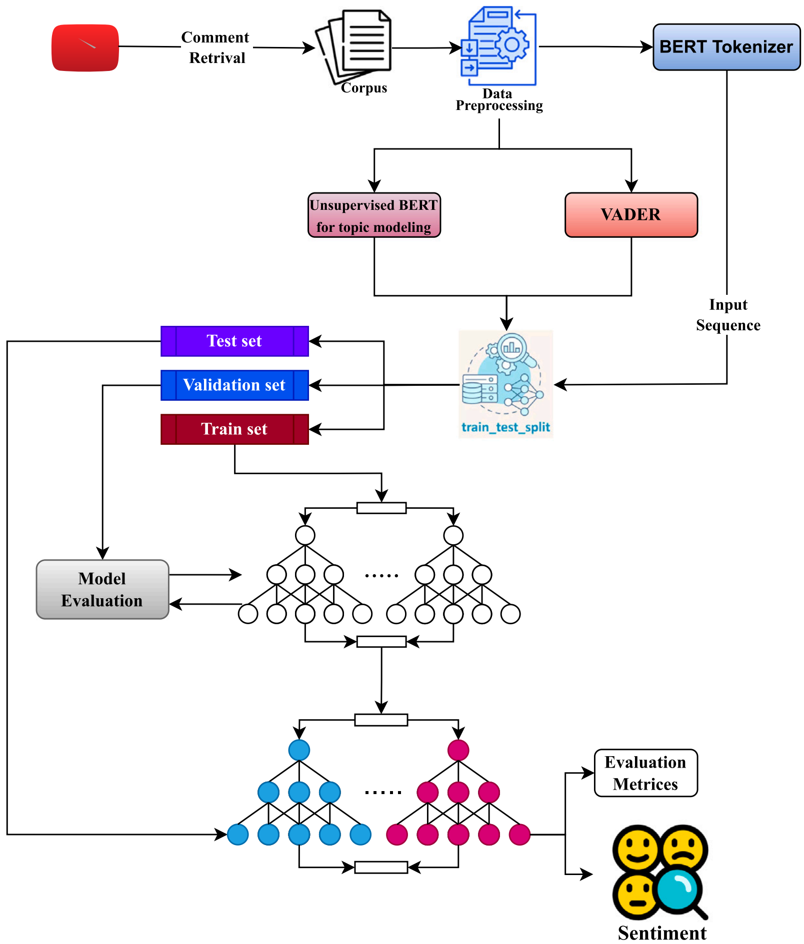

Fig. 3. Workflow of the proposed model.

meanings within the sentence (Aurpa et al., 2021). This bidirectional understanding is achieved by the transformer based architecture of BERT (Aurpa, Rifat, Ahmed, Anwar, & Ali, 2022), which allows self-attention mechanisms to weight the importance of words in relation to each other, independent of the position in the sentence.

Pre-training BERT : BERT (Bidirectional Encoder Representations from Transformers) is pre-trained on large corpora using two tasks: Masked Language Modeling (MLM), where a portion of words is masked, and the model predicts them using context and Next Sentence Prediction (NSP), where the model learns sentence relationships. These tasks enable BERT to capture bidirectional context and sentence semantics.

Fine-tuning BERT : Fine-tuning of heads involves adapting pre-trained BERT into tasks at hand by adding the task-specific layer and training on labeled data. This is an efficient process that updates both pre-trained weights and the new layer, with a minimum of data arriving at task-specific optimization.

This has facilitated several modified architecture developments with a view to performance improvement (Gillioz, Casas, Mugellini, & Abou Khaled, 2020). This architecture performs classification of a dataset into separate topics semantically similar.

In this study, clustering techniques, supported by dimensionality reduction with UMAP, are used to classify datasets into semantically similar topics. UMAP not only reduces computational overhead but also helps preserve the intrinsic structure of data clusters, enabling better separation of topics (George & Sumathy, 2023). Here is clustering Eq. (1)

$$\underset{C}{\text{minimize}} \sum_{i=1}^{K} \sum_{x \in C_i} \|x - \mu_i\|^2 \tag{1}$$

where:

- K is the number of clusters,
- C_i represents the set of points assigned to cluster i,
- $x \in C_i$ is a data point in cluster i,
- μ_i is the centroid of cluster i, calculated as:

The equation of centroid (2) following

$$\mu_i = \frac{1}{|C_i|} \sum_{x \in C_i} x \tag{2}$$

Here, $\|x - \mu_i\|^2$ represents the squared Euclidean distance between a data point x and the centroid μ_i.

3.4.2. CNN

Convolutional Neural Networks (CNNs) are well-known for their ability to capture spatial hierarchies in data, a feature that makes them highly effective for sentence modeling as well. CNN semantic extraction layer consists of convolutional layer and pooling layer (Dong, He, Guo, & Zhang, 2020), In the context of natural language processing (NLP), CNNs apply convolutional filters over word embeddings as well as allowing them to detect local features such as phrases or word combinations that carry specific meanings or semantic cues. Convolutional Neural Networks (CNNs) have demonstrated their effectiveness as powerful semantic composition models for sentence modeling. A typical CNN consists of multiple convolutional and pooling layers, followed by a linear or non-linear classifier (Er, Zhang, Wang, & Pratama, 2016).

The Rectified Linear Unit (ReLU) is one of the simplest and most widely used nonlinear activation functions, achieving notable performance in neural network learning (Alkhouly, Mohammed, & Hefny, 2021).

The convolutional layer decreases the dimension of the node embeddings by creating feature vectors z_i that are determined by sliding a filter $F \in \mathbb{R}^{f \times d_h}$ of length f from i to $i+f-1$ and extracting important information (Phan, Nguyen, & Hwang, 2022) as follows in Eq. (3), (4)

$$z_i = \mathrm{ReLU}\left(F \ominus E_{i:i+|V|-1} + b\right) \tag{3}$$

where $i = [1, |V|]$ is the order of the node representations E; $\ominus$ is the convolution operator; ReLU is an activation function. b is a bias term. Therefore, the feature vectors are created from node representations as follows:

$$z = [z_1, z_2, \ldots, z_{|V|}] \tag{4}$$

Max pooling is a technique used within Convolutional Neural Networks, by which the dimensionalities of feature maps are reduced while preserving the contact of important information. It will be an effective method in our project, allowing us to focus on the most informative patterns or words in every feature map, thus allowing our model to detect key semantic elements without being sensitive to their positions.

The max-pooling layer creates feature vectors of the same size by selecting the maximum number from each vector z_i. The main reason is that the size of the feature vectors z_i depends on the dimensions of both matrices E and F. Therefore, the dimensions of vectors $z_i \in z$ will differ if the sentence length and filter size are different. New feature vectors $\hat{z}$ are defined as follows in Eq. (5) and (6).

$$\hat{z} = [\hat{z}_1, \hat{z}_2, \ldots, \hat{z}_{|V|}] \tag{5}$$

where

$$\hat{z}_i = \mathrm{Max}(z_i) \tag{6}$$

The fully connected layer fine-tunes the characteristics of the previous layers to determine the aspect sentiment as follows in Eq. (7):

$$\hat{y} = \mathrm{Softmax}(W^E \cdot \hat{z} + b) \tag{7}$$

where $W^E \in \mathbb{R}^{l \times |V|}$ and $b \in \mathbb{R}^l$ are a weight matrix and a bias of this layer. l is the number of sentiment classes.

3.5. Proposed classifier

In this research, we have proposed a BERT-CNN architecture for the topic-wise sentiment classification. Unlike existing works, we intend to pass the topic as an individual feature besides sending the text. The Fig. 2 illustrates the proposed model. Different layers used in the figure are given below:

- **BERT Tokenizer :** Breaks the input text into tokens while adding special tokens like [CLS] and [SEP]. This ensures the input adheres to BERT's format and retains semantic meaning. Provide three different input sequences: token embeddings, segment embeddings, and position encoding.
- **BERT Layer :** The sequence generated by the tokenizer is sent to this layer. Leverages self-attention mechanisms and positional encoding to capture contextual relationships between tokens. This BERT layer allows the model to understand the context of words in a text.
- **1D Convolutional Layer :** Extracts local features from the output of the transformer layers. This layer focuses on identifying patterns or specific n-grams in the data, complementing the contextual information provided by BERT.
- **Global Max Pooling :** Reduces the dimensionality of the feature maps by taking the most important features (maximum values). This step ensures that the most critical local features are retained for further processing.
- **Concatenation Layer :** Merges the features extracted from the convolutional layer with a fully connected layer's output that processes topic information. This integration allows the model to combine sentiment and topic-specific features effectively.
- **Fully Connected Layer :** Processes the concatenated features to create a high-level representation for the final classification task. This layer contributes to adjusting the dimension of the text feature and numeric features' dimensions while merging them.
- **Softmax Layer :** Converts the output of the fully connected layer into a probability distribution across sentiment classes (e.g., positive, negative, neutral) using the softmax function.

In Fig. 2, at first, we send the text through the BERT tokenizer and get three different vectors: Token embeddings, Positional Encoding, and Segment embedding. These three embeddings are passed to a BERT layer and followed by a 1D Convolutional layer. The inclusion of CNN plays crucial role in capturing text patterns. The input reaches the Global Max Pooling layer. CNNs allow parallel processing and reduce computational complexity while maintaining strong feature extraction capabilities. Parallelly, the topic is encoded as a numeric value and sent to a fully connected layer. This fully connected layer plays a crucial role in ensuring that the numeric feature matches the dimension of the text features, providing a solid foundation for our sentiment classification process. Finally, the text feature(output of the Global Max Pooling layer) and the numeric features are combined in a Concatenation Layer. This step is significant as it prepares the data for the final stage of prediction. The model then creates the prediction through another fully connected layer and Softmax, providing a comprehensive understanding of the sentiment classification process.

By leveraging BERT-CNN within BERT architecture, our model effectively captures both contextual dependencies and fine gained text patterns, leading to improve our model performance.

3.6. Model evaluation

For the evaluation of our Model, We have utilized a variety of evaluation metrics like Accuracy, Precision, F1 Score, Macro Avg and Weighten Avg. The calculation of Confusion Matrix has played a very important role for our observation. Here is the following equations are:

1. **Accuracy** measures the proportion of correctly predicted instances to the total number of instances using the following Eq. (8):

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Instances}} \tag{8}$$

2. **Precision** quantifies the accuracy of positive predictions for each class through this Eq. (9):

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{9}$$

3. **Recall (Sensitivity)** evaluates the ability of the model to identify all positive instances for each class by this Eq. (10):

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{10}$$

4. **F1-Score** is the harmonic mean of Precision and Recall, providing a balanced measure with the following Eq. (11):

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{11}$$

3.7. Workflow of proposed framework

Fig. 3 represents the workflow diagram of the proposed methodology. Here the steps are explained below:

1. **Comment Retrieval:** First, we collected comments from popular news channels, such as CNN and BBC's YouTube videos.
2. **Data Preprocessing:** The collected text is cleaned and prepared for analysis. Different types of data preprocessing are conducted to convert the comments into helpful corpus.

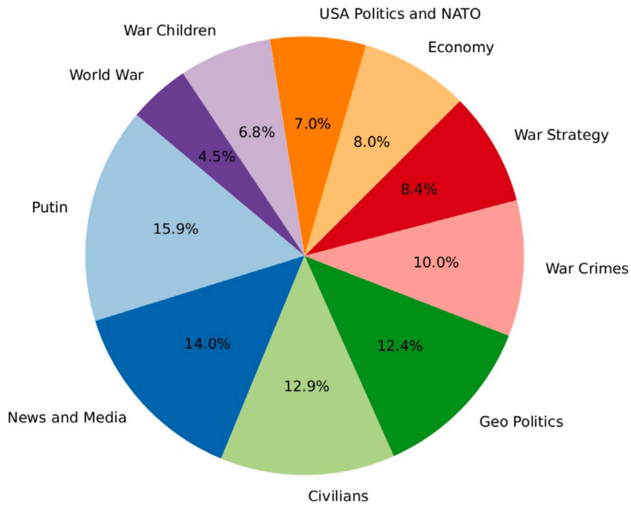

Fig. 4. Pie graph of the Russia-Ukraine War and the latent topics of dataset.

3. **Topic Modeling:** Next, we have used an unsupervised BERT model to determine the topic name. This model organizes the comments into groups based on their themes (e.g., politics and humanitarian issues).
4. **Sentiment Analysis:** The sentiment tool used here is VADER. The VADER tool analyzes the comments to determine their sentiment-whether they are positive, negative or neutral.
5. **Model Training:** Now, the preprocessed and labeled data are utilized to train the proposed hybrid model. We split the dataset into three parts: test train and validation dataset. Train and validation datasets are used to train the model.
6. **Model Evaluation:** Then, the test dataset was used to evaluate the model and determine the confusion matrix and other evaluation metrics. Moreover, it is able to generate outputs for new unseen observations.

4. Result

4.1. Performance analysis of the proposed methodology

In this section, we have examined how well our proposed method works. The next parts explain the results of the techniques we used.

4.1.1. Proportion of total comments by topics

This pie chart in Fig. 4 represents the share of the total comments by topics from sentiment analysis of topics related to the Russia-Ukraine War. In this chart, different segments represent a different thematic topics, thereby capturing many faces of discourse related to the conflict. Each slice is of a size proportional to the number of comments associated with a particular topic, highlighting high and low levels of public interest.

Here is an overview of each Topic:

- **Putin**: The comments in this topic are, to a large extent dominated, at 15.9%, representing 13,672 comments. There is great discussion centered around Vladimir Putin, his actions and involvement regarding this conflict.
- **News and Media**: This topic comprises 14.0% of the comments, representing 12,008 comments. Great interest by the public has been recorded with regard to the role that the media has played in reporting and sometimes shaping-narratives about this war.
- **Civilians**: This topics representing 12.9% of comments (11,065 comments) relevant to civilian impact, casualties and concerns about human rights. In addition, this topic shows the empathy of citizens toward people who have been directly affected by the war.
- **Geo Politics**: Representing 12.4% of comments (10,662 comments), this topic covers geopolitics, including alliances, tensions and larger international implications.
- **War Crimes**: This topic includes 10.0% of the comments (8594 comments), focused on accusations or reports of war crimes; hence, this topic emphasizes ethical and legal violations within the conflict.
- **War Strategy**: This topic includes 8.4% of the comments (7220 comments), involving strategic and tactical remarks about the war. It reflects interest in military operations and their consequences.
- **Economy**: Represented with 8.0% of comments (6877 comments) that focuses on economic consequences including sanctions, global market impacts and financial instability in the context of the conflict.
- **USA Politics and NATO**: The topic covers comments with 7.0% (6052 comments) regarding the involvement and instance of the USA and NATO representing some elements of global political dynamics that influences public opinion.
- **War Children**: This topic shows public concern for the impact of the conflict on children at 6.8% or 5866 comments highlighting significant empathy for the younger victims of war.
- **World War (WW2 and WW3)**: This topic at 4.5% (3888 comments), involves discussions that relate the present conflict to World War II or hypothetical World War III scenarios, reflecting the intensity and gravity with which the conflict is viewed.

4.1.2. Explaining the sentiment through WordCloud

Word clouds for a series of the most commonly used words in comments on the Russia-Ukraine War by sentiment: positive, neutral and negative. Each word cloud is in face form to denote the emotion depicted by the sentiment shown in Fig. 5.

Positive Sentiment: This positive word cloud takes the image of a smiling face; it represents words such as optimism, hope, and calls for peace. Words that connote optimism include “like”, “good”, “god”, “want”, “peace”, “great”, “love” and “help”. There are messages of support for all positive events, a longing to move on and even the use of faith or spirituality, as evidenced by “god” and “peace”. These words probably are representative of the people who want to end the war and for collaboration between nations.

Neutral Sentiment: The middle cloud is neutral, in the shape of a face, and contains those words representing less emotional discussions, more factual or opinion discussions of the war. Terms such as “wants”, “media”, “see”, “think”, “moved”, “bbc”, “new”, “know” and “news” indicate interest in the flow of news and opinions on various strategic or political action taken and general observations. “USA” and “Trump” would likely speak to the geopolitical implications of the war and involvement or stance of the United States. There were many people who showed there voice for the peace and avoid the conflicts.

Negative Sentiment: The negative word cloud on the right is shaped like a frowning face. It shows negative words with feelings of anger, sadness, and concern about the conflict. It implies that words like “stop”, “time”, “sad”, “evil”, “hell”, “invade”, “sanction”, “fight” and “military” reflect frustrations with the continued violence and a desire for it to cease. Words like “weapons” and “west” might indicate more critical or fearful feelings about Western countries’ military involvement, while “propaganda” and “killing” might indicate negative perceptions about the war’s narrative or resultant casualties.

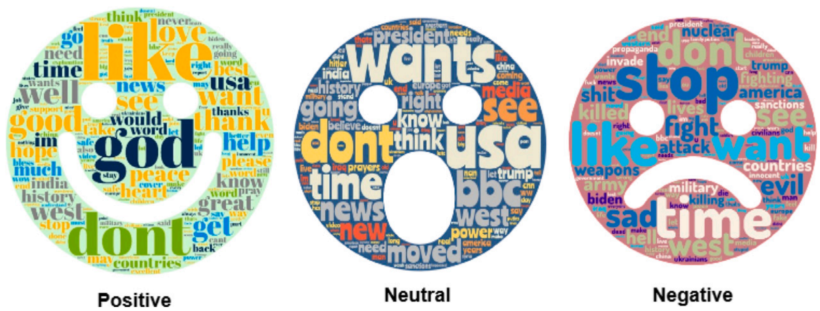

Fig. 5. WordCloud on the Russia-Ukraine War.

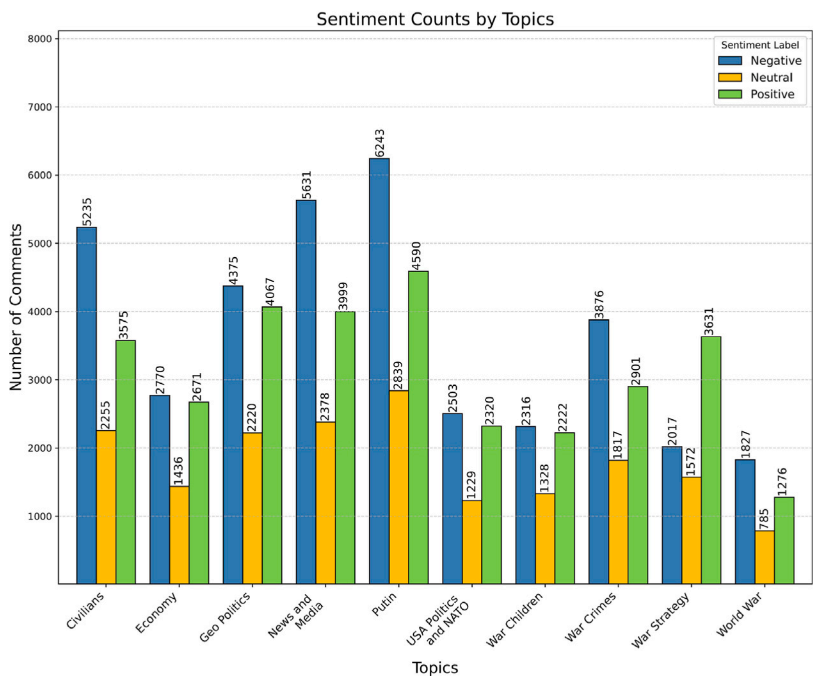

Fig. 6. Topics of the analysis.

4.1.3. Sentiment analysis based on topics:

The bar chart showing the summary of positive, neutral and negative sentiments around the main topics related to the Russia-Ukraine war which are found in the comments of YouTube news like BBC, CNN, DW etc with the News Heading of **Russia launches massive attack across Ukraine | BBC News**, **Fighting with Ukraine in Kursk region enters third day, Russia says | BBC News**, **Expert has theory why Putin has not responded to Ukraine's attack on Russian city**, **See Putin's ominous warning as Russia attacks Ukraine | CNN**, **CNN reporter: This shows just how close Russian forces are to Ukraine capital**, **Fireball after Russian missile hits airport in western Ukraine - BBC News** etc. Each of these topics reflects a thematic causes derived from public discourse and reflects how sentiment varies with subject matter. This distribution points toward areas of concern, criticism and occasionally support or optimism within the global conversation on this conflict. Dominant sentiment patterns reveal emotions, from empathy toward civilians to criticism against political figures and the media. A detailed look at each topic, based on the sentiment distribution, follows in Fig. 6. Here is detail explanation below :

- **Civilians:** The overriding feeling about civilians is that of negative tones. These are represented by 5235 comments of public distress over the humanitarian cost the war exacts on civilian populations. This no doubt covers discussions of casualities, hardships and displacements experienced by everyday people due to the conflict. The 2255 comments for neutral sentiment most likely are fact-based or objective discourse, such as reporting of the events without a strong opinion. Meanwhile, 3575 positive comments may be characterized as empathetic, supportive and admiring of the resilience and bravery of civilians during times of hardship.
- **Economy:** The sentiment about the economy is also predominantly negative, as 2770 comments signal concerns about financial repercussions in terms of inflation, sanctions, hikes in prices and instability in world markets. The neutral sentiment, a consequence of 1436 comments, would express factual discussions on changes in economic indicators. Note that there would be 2671 positive comments to express optimism for economic resilience or expected recovery despite these challenges. This can emanate from discussions on economic adaptation through hope for future stability or confidence in certain regions or industries.
- **Geo Politics:** Geo Politics maintains a close-to-neutral sentiment distribution, having 4375 comments negative charged, 4067 comments positive and 2220 comments neutral. This may be indicative of high variability in opinions on geopolitical alliances and international relations. The discussion points in this topic may include the roles of the United States, the European Union and other nations in regard to Ukraine, be it support or imposing sanctions against Russia. Some of the discussions spill over into the related geopolitical issues of the conflicts of Palestine and

Afghanistan, which shows, indirectly that the war has sharpened interest in global alliances and foreign policies leading to mixed views.

- **News and Media:** The topic News and Media has a highly negative sentiment with 5631 comments, which might represent public criticism against media reporting on the conflict. This could be due to perceived bias, misinformation or even propaganda. It is important to note, however, that positive sentiments or comments are 3999 in number and hence may include factual reports on media coverage. Neutral sentiments, though relatively low, at 2378 comments, may reveal an appreciation or support for what many saw as a balanced or unbiased reporting by credible news sources and underlines the varied response from the general public to media portrayals of the conflict.
- **Putin:** The talk about Putin remains grossly one-sided and negative, with 6243 comments strongly critical or disapproving of his role in this war. This is an indication of the dissatisfaction of the public with actions and decisions made by Putin. However, there are 2839 comments that are classified as neutral, with another 4590 comments being positive, showing mixed and supportive opinions for some groups, possibly because of political affiliations and cultural associations. Since some positive sentiment has been present, it would indicate that some people consider Putin's moves as a show of strength and some form of leadership in the face of global opposition.
- **USA Politics and NATO:** Dominating the topic with negative sentiment, 2503 comments speak to distrust or criticism of Western involvement in the conflict. This may mean skepticism of motives behind NATO involvement and/or criticism of strategies set in place by the West, such as military aid to Ukraine. The neutral comments, 1229 in number, may reflect more objective discussions of the involvement, while the positive comments at 2320 do indeed indicate that some approval or support for Western strategies might emanate from a belief that such involvement is necessary in order to counter Russia's actions. To those supporting them, it is a matter of protecting Ukraine itself and bringing a semblance of stability to the region.
- **War Children:** The theme War Children evokes negativity toward the citizen's feelings, which is reflected in 2316 comments on people's concerns of the war implications on children. It may involve sympathy toward the young victims and accusation toward offenders who commit harm to those innocent lives that should never face such experiences. There are 1328 comments which might be more descriptive or purely factual in nature, hence neutral in tone, as it will talk of the plight of children but not use emotive language. Positive Sentiment 2222, The likely meaning here is support for humanitarian aid efforts, public empathize and calls for international action to protect and assist the children affected by the conflict.
- **War Crimes:** War Crimes are very badly received; for example, 3876 comments condemn the alleged violations and atrocities of the conflict. Sentiment might reflect an outcry from the public over the humanitarian abuses of the war. The neutral comments, 1817 in number, might report on or analyze these incidents objectively. Positive 2901 comments could be considered where justice/accountability-related measures are being called upon, like prosecuting people guilty of war crimes to indicate justice and resolution of some sort.
- **War Strategy:** This discussion thread is overwhelmingly positive, featuring 3631 comments complimenting Russia's military tactics and Ukraine's defense strategy. These comments would be those that praised an individual move or decision in context, thought to have been particularly effective or impressive within the larger context of the war. Neutral comments total 1572, probably featuring more analytical approaches to military strategy with a factual, non-emotional assessment. Meanwhile, 2017 negative comments express disapproval of certain tactics or overall strategies. Also, it shows how people facing the problems for the war situations, people were killed and becoming homeless etc. Imposing people to the war.
- **World War:** World War discussions are characterized by high negative sentiment, at 1827 comments, due to fears of the current conflict scaling into a global war. Comments rated neutral at 785 may include historical comparisons or analysis of previous world wars. Positive sentiment at 1287, is more unusual but can reflect discussions around lessons to be learned from past conflicts that show a focus on lessons learned in adversity rather than celebration of the current crisis. Try to avoid the war.

Table 2
Performance Metrics for Different Datasets.

Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
YouTube	92.26	92.31	91.59	91.91
Reddit	88.87	89.14	89.04	89.03
Twitter	92.98	93.03	92.39	92.64

4.1.4. Performance analysis of model across different dataset

Here Fig. 7, in this figure, three HeatMap is shown based on the Confusion Matrix. The heatmap is capable of accurately describing the relationship of the sentiments. The graphs provides a detailed breakdown of model performances by comparing the predicted and actual labels of the sentiments. We have applied our model across different dataset to figure out our model capability, We have applied the model on primary dataset and beside it, we have applied the model on Twitter dataset and Reddit dataset. From this we have found 88.87% accuracy on Reddit dataset. Reddit Dataset was consisting 64906 texts. Where we found accuracy of 92.98% on Twitter Dataset, The dataset was consisting 17951 texts. Where our dataset got 92.26% of accuracy consisting 85904 comments. The following Fig. 7(Performance Analysis). The model predicted the 7037 instances as True Negative, 3067 as True Neutral and 5630 as True Positive instances from our Primary Dataset(YouTube). In Twitter Dataset we have seen 1000 instances were predicted as True Negative, 867 instances as True Neutral and 637 instances as True Positive.

The sentiment classification model has some weaknesses across all three datasets, with the most errors showing up in the Reddit dataset. Starting with YouTube, while it performs better, there are still some errors. For example, 221 neutral comments and 103 positive ones were wrongly classified as negative. Moreover, 452 negative comments were misclassified as positive. As the YouTube dataset is the biggest dataset, the error number is larger here. The reason of misclassification might me words play, where some negative words were used but the sentence was neutral, so the method predicted it as negative instead of neutral. Twitter, which is a smaller dataset than the remaining two, has shown better performance than other datasets. For instance, 30 negative and 31 positive tweets were misclassified as neutral. This shows that the model has trouble figuring out tweets with middle-ground sentiments and might be oversimplifying the language. The reason of misclassification can be the short tweets, are always tough to identify the sentiments. The most wrong predictions come from the Reddit dataset. Here, 371 positive comments were predicted as the negative class and again, 205 negative comments were wrongly classified as positive. This shows that the model really struggles with Reddit's informal and nuanced conversations.

The overall errors indicates the sentiments are challenging to identify mostly among neutral and negative. Additionally informal languages, sarcasm, short texts contribute to misclassification.

Additionally, We applied our model on Reddit Dataset where we found 2777 instances were predicted as True Negative, 2433 instances as True Neutral and 2668 instances as True Positive.

Table 2 summarizes the model's performance on three dataset on various evaluation metrics: Precision, Recall, F1-Score and Accuracy.

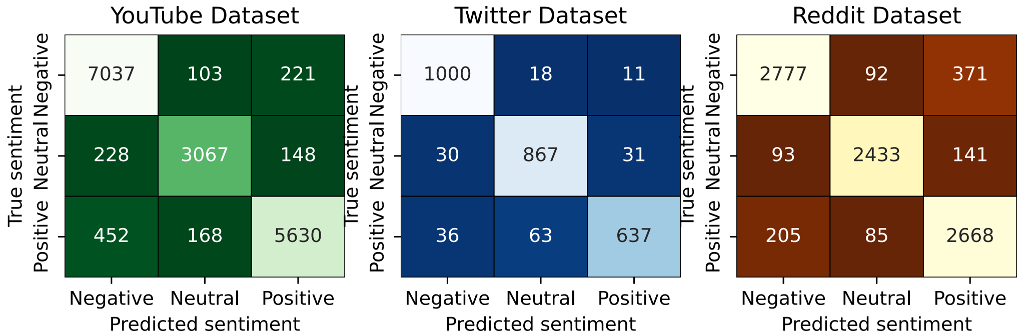

Fig. 7. Performance Analysis.

Table 3
Sentiment Distribution by Model.

Sentiment category	VADER count	DistilBERT count	LSTM count	RoBERTa count
Negative	36,793	60,268	85,904	40,785
Positive	31,252	25,420	0	10,122
Neutral	17,859	216	0	34,997

- YouTube Dataset (Primary) : The model achieved an impressive Accuracy of 92.26%, Precision of 92.31%, Recall 91.59% and 91.91% F1 Score. This indicates the model performed very well.
- Reddit Dataset : The Model Achieved 88.87% Accuracy, 89.14% Precision, 89.04% Recall and 89.03% F1-Score. In this dataset performed slightly lower than the Primary Dataset. But the overall was good enough.
- Twitter Dataset : The model achieved 92.98% Accuracy on Twitter Dataset, 93.03% Precision, 92.39% Recall and 92.64% F1 Score. The model performed exceptionally well on Twitter Dataset.

From Fig. 7 and Table 2 we have seen our model have performed very well across different dataset. Indicating our proposed model is robust and reliable.

Though the model's performance is remarkable for all the datasets we have used here, the model shows the best performance with the YouTube dataset. The YouTube dataset contains a larger amount of data than the other two datasets, which is beneficial for properly training such a heavy model.

4.2. Comparison of the proposed methodology

4.2.1. Comparison of different sentiment labeling

In addition to our proposed model, we have come with various sentiment labeling techniques and topic modeling approaches. Below is a comparison of sentiment classification results across different methods, including VADER, DistilBERT, LSTM and RoBERTa.

Table 3 illustrates the sentiment distribution obtained from these models. The results highlight significant variations in how each model classifies sentiment, particularly in the neutral and positive categories. During our analysis and manual evaluation mentioned in the paper, We found VADER classifies the sentiments in comments more accurately compared to other techniques which were applied. According to our Humanized Evaluation, We picked different samples from the dataset and checked the corresponding sentiment labels: VADER Sentiment Labels, DistilBERT, LSTM and RoBERTa Labels. With our evaluation, we found that VADER is better than other sentiment labelings. The remaining techniques were not suitable for providing a proper sentiment distribution for conducting further experiments. Therefore, we have to use VADER as our sentiment tool.

4.2.2. Comparison of different topic model

Here in Fig. 8, We have shown a comparison of our BERT based Unsupervised Topic Modeling and the LDA Topic Modeling.

By applying LDA we have found topics like : **Putin and Global Threats**, **Civilian and Opinions**, **Media and Propaganda**, **International Relations and Diplomacy**, **Geopolitics and Sanctions**, **US-Russia Political Dynamics**, **Civilians and War Impact**, **Middle East and Defense Strategies**, **Military and Nuclear Concerns** and **Humanitarian Crisis and Peace Appeals** . Here are some topics are similar to the topics we have found in our BERT-based approach. Like : LDA topic "Putin and Global Threats and topic from BERT-CNN approach "Putin" is similar, in both topic Putin were cursed by people, some praised is tactics and many people compared him with other dictators. Another LDA topic "Civilians and Opinion" and BERT-CNN topic "Civilians", these two topics were having same kind of comments, Here the impact were discussed, people shared there opinion on the war. LDA topic "Geo Politics and Sanctions" and BERT-CNN topic "Geo Politics" also same, as well as the LDA topic "US-Russia Political Dynamics" and BERT-CNN topic "USA Politics and NATO" related comments were shared among people. LDA topic "Media and Propaganda" and BERT-CNN topic "News and Media" also matched in some of the portions. LDA topic "Humanitarian Crisis and Peace Appeals" and BERT-CNN topic "War Crimes" and "War Strategy" were got matched.

The Fig. 8 explains that the performance of both models was evaluated using F1-score and Accuracy. The comparison shows the superior performance of the BERT-based model, which achieves higher scores in all metrics.

- F1-score: The BERT-based technique achieves 91.91% and LDA achieves 88.12%, which indicates that BERT is better in generalizing different topics.
- Accuracy: The BERT-based model achieves a total accuracy of 92.26%, while LDA lags behind at 88.67%, reinforcing the effectiveness of contextual embeddings in topic modeling.
  The graph in Fig. 8 clearly shows the significant improvement achieved by the unsupervised BERT-based topic modeling. It highlights that our BERT based model handle the task better than Regular LDA topic modeling by making topics more specific, with enhanced contextual comprehension and higher classification precision.

4.2.3. Comparison of different architecture

Fig. 9 is a comparative evaluation of different transformer-based architecture, such as BERT, XLNet, RoBERTa and ELECTRA with our

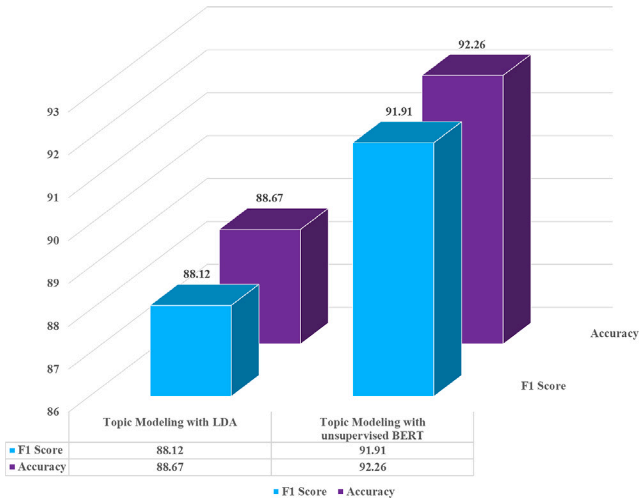

Fig. 8. Comparison of Different Topic Modeling.

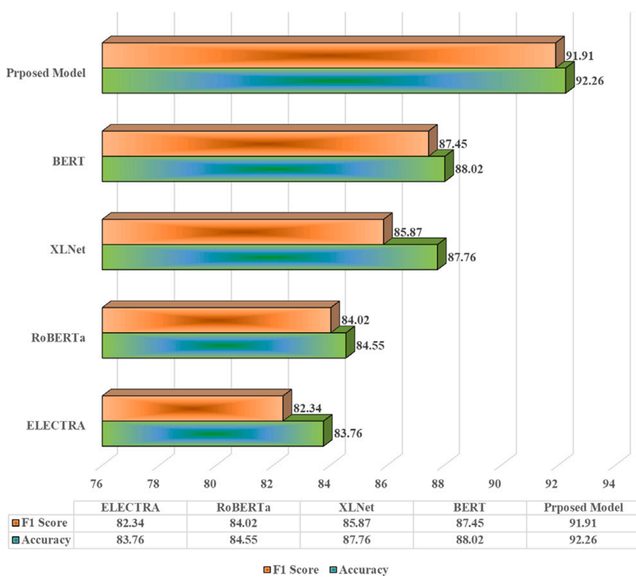

Fig. 9. Comparison of Different Topic Modeling.

proposed model. To align the transformer models with our task, we concrete a transformer layer (for tokenized text input) with a numeric layer (for topic input) similar to our proposed hybrid architecture. The models were compared based on key performance metrics: F1-score and Accuracy. The results highlights the effectiveness of our Proposed Model, which performed well comparing to other architectures across all metrics.

- Proposed Model : Achieves the highest F1-score (91.91%) and Accuracy (92.26%), reflecting its superior ability to understand contextual relationships in contexts and generate more related topics.
- BERT : Shows F1-score (87.45%) and Accuracy (88.02%). While BERT is effective in understanding word relationships (George & Sumathy, 2023), it struggles with long-range dependencies compared to more advanced models. It provided second highest scores in all metrics.
- XLNet : It comes with F1-score (85.87%) and Accuracy (87.76%), as its permutation-based training allows for better bidirectional context learning (Koufakou, 2024). It was less than BERT performance.
- RoBERTa : Achieves F1-score (84.02%) and Accuracy (84.55%), benefiting from dynamic masking and larger training data, But still could not able to manage the better scores in evaluation metrics.
- ELECTRA : Shows the lowest performance with F1-score (82.34%) and Accuracy (83.76%). Despite its efficiency in training via replaced token detection, it lacks the strong contextual learning ability required for topic modeling.

The graphical representation in Fig. 9 clearly shows that our proposed model has performed well comparing to the other existing models. It shows our model is more effective to find the topic, there relation with the texts and good accuracy in term of topics interpretation. Comparing the other transformer models, our proposed BERT-CNN model outperformed. BERT is popular for capturing the contextual pattern of any language by its pretraining task, and CNN enhances the critical features. Combining BERT's deep contextual embeddings with CNN's ability to focus on local patterns, the hybrid model can achieve superior results compared to other transformer models.

5. Discussion

In our work, we had collected YouTube comments from various international media to find out the thematic trends on the Russia-Ukraine war. We also collected data from kaggle for further analysis of our model. We have applied BERT based Unsupervised Clustering for Topic Modeling, VADER For knowing the sentence labeling and there sentiments on the comments, By doing so, We have found many topics but the significant topics mentioned in Fig. 4(Pie-Graph) were **Putin** as Dictator for his attack on Russia, We found a topic as **Economy** which shows the impact of the Russia-Ukraine War, the upsurge of the daily commodities and oil price hike. The output shows a huge number of comments on **Geo Politics** and Bias behavior of NATO and USA on the Russia-Ukraine War. We have seen the role of **News and Media** on the Russia-Ukraine War. We found some comments on how the world is moving for another **world-War**. There are some comments on **War-Crimes**, **War-Strategy** another significant topic was the impact of the in **Civilians**. How they lost lives of there close one, How to be feel when you are homeless. The attack has impact on children we found it as **War-Children**, Where some of the comments were heart breaking that children were killed due to missile attacks, School, College were demolished.

When we go for the WordCloud in Fig. 5 based on Positive, Neutral and Negative sentiment labeling. We have seen words Like, god, help, thank, usa, please which signifies that people were looking for help for the war affected country, looking for assistance for war affected. In Negative Word Cloud we saw words like Wants, USA, BBC, News, Right, trump, president the words were signifies role of the news media and there reporting, again some words related USA found which reflects political dynamics and international involvement. Here some thematic trends came regarding the USA-Election. In negative cloud we had seen some negative worlds like kill, army, attack, west, evil etc which shoes the evilness of war. We have seen the number of positive, negative and neutral sentiments on different topics found from the comments. We have also generated a heatmap of confusion matrix in Fig. 7 based, Our model has correctly predicted 7037 instances as true negative, 3067 instances as true neutral and 5630 instances as true positive in primary dataset.

In our BERT-CNN Hybrid Model, Where We found accuracy of 92.26% in YouTube Dataset, 88.87% of accuracy in Reddit Dataset and 92.98% Accuracy in Twitter Dataset as mentioned in Table 2, Which confirms the effectiveness of our model classifying the sentiments of YouTube Comments. Here CNN were used to feature filtering of our text (Mao, Zhang, & Guan, 2021). CNN extract higher-level abstract features that capture contextual and semantic relationships between words. CNNs concentrate on capturing the essential meaning, which helps reduce the impact of feature noise, even though they have a wider scope (Jacovi, Shalom, & Goldberg, 2018; Zhou, Li, Chi, Tang, & Zheng, 2022). Furthermore, CNNs contribute to the model's performance by identifying distinct topic distributions, which improved the performance and accuracy (She & Zhang, 2018) of our BERT-CNN Hybrid Model.

We have do comparison based on Sentiment Labeling, Topic Modeling and Different Architecture. From sentiment labeling we have seen our VADER labeling was better comparing to the DistilBERT, LSTM and RoBERTa. For justification we also do humanize evaluation of labeling. We noticed that VADER was performed well comparing to other Labeling Techniques.

From the Comparison of Topic Modeling, we have seen that BERT-CNN hybrid Model acquired higher accuracy than LDA. Here BERT based method obtained 92.26% and LDA achieved 88.67% accuracy. In comparing of topics, BERT given topics were more relative to the dataset.

For Further evaluation of Model, we have compared with different architectures, Here also our Proposed Model outperformed gaining highest accuracy of 92.26% comparing to BERT (88.02% accuracy), XLNet (87.76% accuracy) and RoBERTa (84.55% accuracy).

By combining the BERT based Unsupervised Clustering, CNN and VADER for topics extractions and sentiment labeling. Our Hybrid model shows the robustness and superior performance. The accuracy of our model along with our study validates effectiveness of our model. Our research reveals various aspects of the ongoing war crisis, shedding light on global positions and relationships between nations and their allies.

6. Conclusion

We analyzed YouTube comments from various news channels to understand public opinions and trends related to the Russia-Ukraine war. Using a BERT-CNN hybrid model, we identified ten key topics, such as economic concerns, geopolitical issues, humanitarian crises, war strategies, and the role of the media. The comments were categorized into positive, negative, and neutral sentiments, offering a clear picture of public views. Our BERT-CNN model achieved an impressive accuracy of 92.26%, combining BERT's ability to understand context with CNN's feature extraction capabilities. This made it particularly effective in analyzing the complex nature of YouTube comments.

For justifying our analysis and model we have go through various comparisons: Sentiment Labeling Comparison, Topic Modeling Technique Comparison and Model Architecture Comparison, In every comparison our proposed model shows it significant effectiveness and accuracy.

Even in performance analysis our model was applied in two more dataset (Reddit and Twitter) beside primary dataset. The model's robustness was evident in its accuracy. The model obtained 92.26% accuracy on YouTube Dataset, 88.87% accuracy in Reddit Dataset and 92.98% accuracy in Twitter Dataset.

This highlights the potential of using unsupervised clustering and sentiment analysis for analyzing large-scale public discourse.

In conclusion, our approach shows that combining advanced methods can effectively analyze public opinions on global issues like the Russia-Ukraine war. Expanding data sources and including multilingual analysis in future research could offer a more complete understanding of worldwide sentiments and perspectives.

6.1. Limitations

- Language Bias : Our analysis and model both focused on English Language, Sentiments by other language were overlooked.
- Model Complexity and Time Consumption : The Hybrid BERT-CNN model was complex and time consuming for training and fine-tuning.

6.2. Future works

- Aspect based Sentiment Analysis : Future Research would be aspect based, where we will find the sentiments related to the aspects of war, humanitarian and politics.
- Multilingual Data Expansion : Incorporating non-English comments to capture a broader and global perspective of war.
- Model Optimization : Simplifying the hybrid model and optimizing training process for faster execution.
- Context Aware Sentiments Models : The model may understand the sarcasm, short-texts sentiments better.

7. Data collection and privacy

For this research, we made sure not to invade anyone's privacy. We only used comments that are already publicly available. We followed ethical guidelines, ensuring that everyone's privacy was respected and no private information was used.

CRediT authorship contribution statement

Md. Saiful Islam: Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Mahmuda Ferdusi:** Writing – original draft, Writing – review & editing. **Tanjim Taharat Aurpa:** Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Ahmed, Md Shoaib, Aurpa, Tanjim Taharat, & Anwar, Md Musfique (2020). Query oriented topical clusters detection for top-k trending topics in twitter. In *2020 IEEE 8th R10 humanitarian technology conference* (pp. 1–6). IEEE.

Al Maruf, Abdullah, Ziyad, Zakaria Masud, Haque, Md Mahmudul, & Khanam, Fahima (2022). Emotion detection from text and sentiment analysis of Ukraine Russia war using machine learning technique. *International Journal of Advanced Computer Science and Applications (IJACSA)*, *13*(12).

Alkhouly, Asmaa A., Mohammed, Ammar, & Hefny, Hesham A. (2021). Improving the performance of deep neural networks using two proposed activation functions. *IEEE Access*, *9*, 82249–82271.

Aslan, Serpil (2023). A deep learning-based sentiment analysis approach (MF-CNN-BILSTM) and topic modeling of tweets related to the Ukraine–Russia conflict. *Applied Soft Computing*, *143*, Article 110404.

Assaf, Rima, Gupta, Deeksha, & Kumar, Rahul (2023). The price of war: Effect of the Russia–Ukraine war on the global financial market. *The Journal of Economic Asymmetries*, *28*, Article e00328.

Aurpa, Tanjim Taharat, & Ahmed, Md Shoaib (2024). An ensemble novel architecture for Bangla Mathematical Entity Recognition (MER) using transformer based learning. *Heliyon*, *10*(3).

Aurpa, Tanjim Taharat, Ahmed, Md Shoaib, Sadik, Rifat, Anwar, Sabbir, Adnan, Md Abdul Mazid, & Anwar, Md Musfique (2021). Progressive guidance categorization using transformer-based deep neural network architecture. In *International conference on hybrid intelligent systems* (pp. 344–353). Springer.

Aurpa, Tanjim Taharat, Rifat, Richita Khandakar, Ahmed, Md Shoaib, Anwar, Md Musfique, & Ali, A. B. M. Shawkat (2022). Reading comprehension based question answering system in Bangla language with transformer-based learning. *Heliyon*, *8*(10).

Bauskar, Shubham, Badole, Vijay, Jain, Prajal, & Chawla, Meenu (2019). Natural language processing based hybrid model for detecting fake news using content-based features and social features. *International Journal of Information Engineering and Electronic Business*, *11*(4), 1–10.

Chauhan, Uttam, Shah, Shrusti, Shiroya, Dharati, Solanki, Dipti, Patel, Zeel, Bhatia, Jitendra, et al. (2023). Modeling topics in DFA-based lemmatized gujarati text. *Sensors*, *23*(5), 2708.

Dean, Matthew C., & Porter, Ben (2024). Sentiment analysis of Russian-language social media posts discussing the 2022 Russian invasion of Ukraine. *Armed Forces & Society*, 0095327X241235987.

Dong, Junchao, He, Feijuan, Guo, Yunchuan, & Zhang, Huibing (2020). A commodity review sentiment analysis based on BERT-CNN model. In *2020 5th international conference on computer and communication systems* (pp. 143–147). IEEE.

Er, Meng Joo, Zhang, Yong, Wang, Ning, & Pratama, Mahardhika (2016). Attention pooling-based convolutional neural network for sentence modelling. *Information Sciences*, *373*, 388–403.

George, Lijimol, & Sumathy, P. (2023). An integrated clustering and BERT framework for improved topic modeling. *International Journal of Information Technology*, *15*(4), 2187–2195.

Gillioz, Anthony, Casas, Jacky, Mugellini, Elena, & Abou Khaled, Omar (2020). Overview of the transformer-based models for NLP tasks. In *2020 15th conference on computer science and information systems* (pp. 179–183). IEEE.

Haddi, Emma, Liu, Xiaohui, & Shi, Yong (2013). The role of text pre-processing in sentiment analysis. *Procedia Computer Science*, *17*, 26–32.

Hasan, Mahmud, Islam, Labiba, Jahan, Ismat, Meem, Sabrina Mannan, & Rahman, Rashedur M. (2023). Natural language processing and sentiment analysis on bangla social media comments on Russia–Ukraine war using transformers. *Vietnam Journal of Computer Science*, *10*(03), 329–356.

Hu, Xia, Tang, Jiliang, Gao, Huiji, & Liu, Huan (2013). Unsupervised sentiment analysis with emotional signals. In *Proceedings of the 22nd international conference on world wide web* (pp. 607–618).

Jacovi, Alon, Shalom, Oren Sar, & Goldberg, Yoav (2018). Understanding convolutional neural networks for text classification. arXiv preprint arXiv:1809.08037.

Jeba, Samiha Maisha, Aurpa, Tanjim Taharat, Siyam, Farshid Hossain, Khan, Rajib, & Mansia, Faniyam Maria (2023). Analysis of public sentiment on dhaka metro rail with transformer based architectures. In *2023 26th international conference on computer and information technology* (pp. 1–6). IEEE.

Koufakou, Anna (2024). Deep learning for opinion mining and topic classification of course reviews. *Education and Information Technologies*, *29*(3), 2973–2997.

Krivičić, Armin, & Martinčić-Ipšić, Sanda (2023). Analyzing sentiment of reddit posts for the russia-ukraine war. In *2023 46th MIPRO ICT and electronics convention* (pp. 1709–1714). IEEE.

Kumarappan, J., Rajasekar, E., Vairavasundaram, S., Kotecha, K., & Kulkarni, A. (2024). Siamese graph convolutional split-attention network with NLP based social sentimental data for enhanced stock price predictions. *Journal of Big Data*, *11*(1), 154.

Liu, Chuchu, Fang, Fan, Lin, Xu, Cai, Tie, Tan, Xu, Liu, Jianguo, et al. (2021). Improving sentiment analysis accuracy with emoji embedding. *Journal of Safety Science and Resilience*, *2*(4), 246–252.

Madhuri, Simhadri, & Lakshmi, S. Venkata (2021). Detecting emotion from natural language text using hybrid and NLP pre-trained models. *Turkish Journal of Computer and Mathematics Education*, *12*(10), 4095–4103.

Maheshwari, Himani, Chandra, Umesh, Yadav, Dharmendra, & Gupta, Ashulekha (2023). Twitter sentiment analysis in the crisis between Russia and Ukraine using the bert and LSTM model. In *2023 international conference on computing, communication, and intelligent systems* (pp. 1153–1158). IEEE.

Mao, Song, Zhang, Lu-Lu, & Guan, Zhen-Guo (2021). An LSTM&Topic-CNN model for classification of online Chinese medical questions. *IEEE Access*, *9*, 52580–52589.

Noor, Shagofah, Tajik, Omid, & Golzar, Jawad (2022). Simple random sampling. *International Journal of Education & Language Studies*, *1*(2), 78–82.

Orhan, Ebru (2022). The effects of the Russia–Ukraine war on global trade. *Journal of International Trade, Logistics and Law*, *8*(1), 141–146.

Pano, Toni, & Kashef, Rasha (2020). A complete VADER-based sentiment analysis of bitcoin (BTC) tweets during the era of COVID-19. *Big Data and Cognitive Computing*, *4*(4), 33.

Pereira, Paulo, Bašić, Ferdo, Bogunovic, Igor, & Barcelo, Damia (2022). Russian-ukrainian war impacts the total environment. *Science of the Total Environment*, *837*, Article 155865.

Phan, Huyen Trang, Nguyen, Ngoc Thanh, & Hwang, Dosam (2022). Aspect-level sentiment analysis using CNN over BERT-GCN. *IEEE Access*, *10*, 110402–110409.

Pietrzak, Piotr (2024). The Russia–Ukraine war and the renaissance of IR realism. In *Forum nauk społecznych* (2), (pp. 9–24). Wydawnictwo Uniwersytetu Warmińsko-Mazurskiego w Olsztynie.

Pradha, Saurav, Halgamuge, Malka N., & Vinh, Nguyen Tran Quoc (2019). Effective text data preprocessing technique for sentiment analysis in social media data. In *2019 11th international conference on knowledge and systems engineering* (pp. 1–8). IEEE.

Pradhan, Rahul (2021). Extracting sentiments from youtube comments. In *2021 sixth international conference on image information processing*: *vol. 6*, (pp. 1–4). IEEE.

Rajasekar, Elakkiya, Chandra, Harshiv, Pears, Nick, Vairavasundaram, Subramaniyaswamy, & Kotecha, Ketan (2025). Lung image quality assessment and diagnosis using generative autoencoders in unsupervised ensemble learning. *Biomedical Signal Processing and Control*, *102*, Article 107268.

Ramos, Leo, & Chang, Oscar (2023). Sentiment analysis of Russia–Ukraine conflict tweets using RoBERTa. *Uniciencia*, *37*(1), 421–431.

Salur, Mehmet Umut, & Aydin, Ilhan (2020). A novel hybrid deep learning model for sentiment classification. *IEEE Access*, *8*, 58080–58093.

Scheerder, DAAN (2024). *Topics and sentiments on the Russo-Ukrainian war* (Ph.D. thesis), Tilburg University.

She, Xiangyang, & Zhang, Di (2018). Text classification based on hybrid CNN-LSTM hybrid model. In *2018 11th international symposium on computational intelligence and design, vol. 2* (pp. 185–189). IEEE.

Shlkamy, Eman Sedqy Ibraihm, Mahar, Khaled Mohammed, & Sedky, Ahmed Ahmed Hesham (2023). A Russia–Ukraine conflict tweets sentiment analysis using bidirectional LSTM network. *International Journal of Science and Research (IJSR)*, *12*(2).

Simarmata, Allwin, Xu, Anthony, Phanie, Matthew Evan, et al. (2023). Sentiment analysis on twitter posts about the Russia and Ukraine war with long short-term memory. *Sinkron: Jurnal Dan Penelitian Teknik Informatika*, *7*(2), 789–797.

Sinha, Soumen, Innani, Saketh, Chinnari, Pawan, & Khan, Mehek (2023). Sentiment analysis of Russia-Ukraine conflict: A hybrid approach using VADER, GloVe-embedding and LSTM. In *Proceedings of the 2023 7th international conference on computer science and artificial intelligence* (pp. 16–22).

Uysal, Alper Kursat, & Gunal, Serkan (2014). The impact of preprocessing on text classification. *Information Processing & Management*, *50*(1), 104–112.

Vaghela, D. B., Makwana, S. H., Chande, H. D., & Mehta, P. (2024). Twitter based sentiment analysis of Russia–Ukraine war using machine learning. *Journal of Electrical Systems*, *20*(10s), 1269–1283.

Vijayarani, S., Ilamathi, Ms J., Nithya, Ms, et al. (2015). Preprocessing techniques for text mining-an overview. *International Journal of Computer Science & Communication Networks*, *5*(1), 7–16.

Wada, Shoya, Takeda, Toshihiro, Okada, Katsuki, Manabe, Shirou, Konishi, Shozo, Kamohara, Jun, et al. (2024). Oversampling effect in pretraining for bidirectional encoder representations from transformers (BERT) to localize medical BERT and enhance biomedical BERT. *Artificial Intelligence in Medicine*, *153*, Article 102889.

Wadhwani, Ganesh Kumar, Varshney, Pankaj Kumar, Gupta, Anjali, & Kumar, Shrawan (2023). Sentiment analysis and comprehensive evaluation of supervised machine learning models using Twitter data on Russia–Ukraine war. *SN Computer Science*, *4*(4), 1–11.

Zhang, Qi, Hu, Yi, Jiao, Jianbin, & Wang, Shouyang (2024). The impact of Russia–Ukraine war on crude oil prices: An EMC framework. *Humanities and Social Sciences Communications*, *11*(1), 1–12.

Zhang, Wenxuan, Li, Xin, Deng, Yang, Bing, Lidong, & Lam, Wai (2022). A survey on aspect-based sentiment analysis: Tasks, methods, and challenges. *IEEE Transactions on Knowledge and Data Engineering*, *35*(11), 11019–11038.

Zhou, Yajian, Li, Jiale, Chi, Junhui, Tang, Wei, & Zheng, Yuqi (2022). Set-CNN: A text convolutional neural network based on semantic extension for short text classification. *Knowledge-Based Systems*, *257*, Article 109948.