Climate Change on Social Media: AI and Deep Learning-Based Analysis of Tweets

Bahar URHAN[1], Mehmet KAYAKUŞ[2], Dilşad ERDOĞAN[3],
Gülten ADALI[4]*, Emrah BOZKURT[5], Zeynep Nihan BAKIR[6]

Department of Public Relations and Promotion-Faculty of Communication, Akdeniz University, Antalya 07058, Türkiye[1]
Department of Biomedical Engineering-Faculty of Engineering, Akdeniz University, Antalya 07058, Türkiye[2]
Department of Finance, Banking and Insurance-Korkuteli Vocational School, Akdeniz University, Antalya 07800, Türkiye[3]
Department of Advertising-Faculty of Communication, Akdeniz University, Antalya 07058, Türkiye[4]
Department of Public Relations and Publicity, Afyon Kocatepe University, Afyonkarahisar 03030, Türkiye[5]
Independent Researcher, Muğla, 48000, Türkiye[6]

***Abstract*—The study analyses Turkish and English tweets about climate change on the social media platform Twitter and comparatively examines individuals" perceptions, concerns, and emotional reactions to this issue. A total of 2,046 Turkish and 18,000 English tweets were collected; 1,104 Turkish and 6,449 English tweets were analyzed after the cleaning process. Artificial intelligence-based methods such as text mining, sentiment analysis, and topic modelling are used. Topic modelling with Latent Dirichlet Allocation (LDA) identified prominent themes in tweets in both languages. Sentiment analysis is performed using deep learning techniques to categories tweets into positive, negative, and neutral categories. The findings show that English tweets contain stronger emotional reactions, while Turkish tweets contain a higher proportion of neutral expressions. Additionally, it was observed that the perception of climate change can differ in local and global contexts. Based on a multidimensional analysis of social media data, the study provides valuable insights into the development of environmental communication strategies. The comparison of Turkish and English tweets contributes to understanding the effects of cultural contexts on climate change perception. The findings have important implications for policymakers and environmental awareness campaigns, as they highlight the need for tailored communication strategies that consider cultural differences in climate change perception.**

***Keywords*—Climate change; awareness; social media; communication; machine learning; sentiment analysis; deep learning**

I. Introduction

Climate change is recognized as one of the greatest environmental, economic, and social challenges of our time. Scientific evidence shows that this phenomenon significantly threatens not only natural systems but also human life and global economic balances [1]. In the 21st century, accelerating climate change due to increasing carbon emissions and greenhouse gases has made social awareness and collective action more important than ever. However, the solution to such a complex and multidimensional problem is closely related not only to scientific developments but also to the attitudes and behaviors of societies shaped through access to information and communication channels. In this context, social media platforms play a central role in educating and raising awareness about climate change.

This study aims to examine how Turkish and English tweets on the social media platform Twitter create a social perception in the context of climate change and how users feel about this issue. The study is based on a dataset consisting of 2046 Turkish and 18,000 English tweets, and after the data cleaning process, 1104 and 6449 tweets were analyzed, respectively. The methods used include text mining, sentiment analysis, and topic modelling, and thus, social media content in different languages is evaluated from a comparative perspective.

In this study, text mining and deep learning methods are used to analyses social media data. Texts were cleaned with natural language processing (NLP) techniques, and n-gram analyses were performed. Latent Dirichlet Allocation (LDA), an unsupervised machine learning method, is used to identify themes in the tweets, allowing for a comparison of themes in different languages. In the sentiment analysis phase, Long Short-Term Memory (LSTM) with a deep learning model is used to classify tweets as positive, negative and neutral. These methods reveal the climate change perceptions and emotional tendencies of social media users in a multidimensional way.

The originality of the study stems from the fact that it comparatively analyses how a global problem such as climate change is perceived at local and global levels through social media data. Although there are many studies in the literature on the impact of social media on environmental awareness, such cross-lingual comparisons are limited. For example, Lewandovsky et al. (2019), while examining how social media affects the perception of climate change, stated that analyses in the local and global contexts are limited [2]. This study aims to fill this gap by revealing how perceptions of climate change differ in different cultural and linguistic contexts through Turkish and English tweets.

Unlike recent multilingual studies that rely on transformer-based architectures such as BERT or XLM-R for cross-lingual alignment, this study adopts a comparative approach using language-specific datasets without explicit embedding alignment. This allows for the exploration of natural linguistic divergence in sentiment and topic structures.

Based on this motivation, the study tests the following hypothesis:

*Corresponding author.

H1: There are statistically significant differences between Turkish and English tweets in terms of sentiment distribution and topic structure.

This research makes three main contributions to literature: Firstly, by providing a multidimensional analysis of social media content on climate change, it facilitates the understanding of individuals' perceptions and attitudes towards the issue. Secondly, the comparative evaluation of social media content in different languages using sentiment analysis and topic modelling methods reveals the effects of cultural contexts. Thirdly, the findings have the potential to provide concrete recommendations for policymakers and stakeholders developing environmental communication strategies.

As the impacts of climate change are rapidly increasing, it is critical to understand individuals' perceptions on this issue and develop the right communication strategies. The results of this study may contribute to raising awareness on combating climate change and shed light on the development of more effective environmental communication strategies through social media.

II. Literature

The last two centuries, during which technology has made it possible to create considerably accurate measures, have not only revealed the phenomenon of climate change but also given us a better understanding of the threat it poses to the well-being of life on Earth and the ecological stability of the planet. These measures have led to a growing international recognition of the threat. In the 20th century, most scientists have not seen greenhouse warming as a severe problem. It took decades of evidence, fierce debates, and hundreds of panels in many countries to convince them they were wrong. In the late 1970s, these efforts led to the conclusion that warming might one day become a problem. But, in the current century, scientific meetings and respected panels have concluded that warming is likely to become a serious problem [3].

Considering current knowledge, climate change is recognized as one of the greatest threats to human life. However, its trajectory can be somewhat affected depending on human activities in the face of this threat [1]. Climate change is predicted to seriously impair the life support systems of many species, drastically reduce the population, and result in crucial and fundamental changes to social structures. Therefore, the importance of communication on the topic is revealed by how to approach it and guarantee cooperation, creative policies, appropriate technological developments, and new behaviors [4]. Pearce et al. (2019), social media such as Twitter are important tools for enabling everyday debate and negotiation on climate change [5]. Dutton argues that the internet acts as a "Fifth Estate" and uses estate theory to illuminate the most significant power shift of the digital age [6]. Through the power of the internet, networked individuals can control facts and make social interventions in a civic context. The internet gives them this power, and digital media essentially increases the information and communication power of ordinary individuals. Communities that can independently and strategically search, create content, network, and collaborate on any topic are able to leak information in this way.

The relationship between social media and the construction of people's ideas about climate change and changing attitudes has been the focus of many studies. For example, Lewandovsky et al. (2019) measured attitudes towards climate change on social media and found that science-based posts reinforced belief in global climate change, but negative comments on these posts reduced perceived consensus [2]. It was observed that social media is a more powerful tool in terms of receiving information and forming attitudes with this information. Accordingly, new media and user-generated content have twice the impact size compared to traditional and centralized content-generating channels [7].

Comparative studies on the relationship between media coverage and climate change, such as our current study, are limited in number but have been able to provide estimates with quite significant data. For example, Thaker's (2023) large-scale comparative study on the coverage of climate change in international media, based on data from the 2022 Yale Meta survey of 110 countries, found that the relationship between perceptions formed through social media and protest intention is related to personal harm. Accordingly, because poor and developing countries are more exposed to the impacts of climate change, they perceive climate change as a greater risk and are more willing to participate in protests, even if such content receives low media coverage [8].

The nature of the media helps shape public opinion on climate change [1]. By their very nature, online posts have several advantages over traditional media in influencing public opinion. Digital activism, which stands out among these, provides an advantage. The most prominent example of this is the case of Greta Thunberg. As seen in the phenomenon of raising online awareness through digital activism, which has entered the literature as the "Greta Effect," Greta Thunberg has influenced both policymakers and the news media to take responsibility for climate change, thanks to her supporters or critics [9].

The existing literature on social media assumes that three main mechanisms work in climate change through posts: norm diffusion, opinion leadership, and public opinion formation [10]. On the other hand, research has found that those who express an opinion on this issue exhibit some categorical reactions: deniers, jokers, advocates, skeptics, and realists [11-14]. Considering such categories in climate communication can better determine its direction. Yet another study found that a high concern for climate resulted in greater engagement in online climate change, and adolescents who cared more about the norms of their peers had higher levels of online climate change because of incidental exposure [15]. In this context, it is seen that many variables, approaches, and categories can be revealed through social media research on climate change. On the other hand, due to the rapid advancement in communication technologies, climate-based information disinformation has become even more possible, and misinformation spread has become extremely easy and fast [2]. Therefore, the fact that a very useful and slippery ground such as social media plays an active role in climate communication emphasizes conscious and effective use as well as some precautions.

Just as Climate Central, which "conveys the science, impacts, and solutions of climate change to the public and decision-makers" (2024), provides reliable information on the issue through its posts on various media channels, supporting social media content with realistic arguments such as visuals and videos, the active participation of scientists, who are seen to have a significant impact on public participation and the responsibility of providing first-hand information, and the role of scientists being free from doubt among both the public and their colleagues can make dialogic and participatory climate communication possible [12].

III. Materials and Methods

In this study, data were collected from the Twitter platform in Turkish and English to analyses social media posts on climate change. The data was collected using the API of the X platform over a period of one month. In total, 2046 Turkish and 18,000 English tweets were obtained. During the data collection process, the keyword "climate change" is used. The focus of the study is to understand the perceptions, concerns, and perspectives of users on this issue. In the text preprocessing stage, repetitive, incomplete, noisy, and irrelevant content was removed, and finally, 1104 Turkish and 6449 English tweets were included in the study.

Formally, the sentiment classification problem can be defined as follows: given a tweet x_i, the goal is to predict its sentiment label $y_i \in$ {positive, negative, neutral}. The model aims to learn a function $f(x_i) \rightarrow y_i$ that minimizes classification error.

The LSTM model was implemented with the following configuration: embedding dimension = 100, batch size = 32, epochs = 10, optimizer = Adam, and dropout rate = 0.5. All experiments were conducted using TensorFlow/Keras, and model evaluation is performed using a 70/30 train-test split.

Text mining, topic modelling, and sentiment analysis methods are used to analyses the tweets. Text mining enabled the analysis of word frequencies and text patterns in tweets, and thus meaningful information was extracted from the data. Topic modelling is used to identify the prominent themes in tweets and to determine which topics users focus on. Sentiment analysis is used to identify positive, negative, and neutral emotional tones in tweets. These methods enabled comparative analysis of social media content in different languages and made it possible to evaluate perceptions on climate change in a multidimensional way.

A. Data Set

The data set used in the study was obtained through the API provided by the X platform. The main reason for choosing the X platform is that it is a social media platform that is heavily used by users and provides data from a wide user base on social issues such as climate change [16]. This increased the inclusiveness of the study and the generalizability of its results.

The keyword "climate change" was preferred in the data collection process. The reason for this is that climate change is of increasing importance in both global and local contexts and that social media users have various opinions and perceptions on this issue, providing a suitable data source for the analyses targeted by the study. In this context, the choice of keywords aims to facilitate the understanding of users' perceptions and perspectives on the issue.

The data was collected over a period of one month using the Python programming language and the API of the X platform. The reason for choosing a one-month time is that this period provides the intensity and diversity of users' posts on the topic that can form a meaningful sample. In addition, this period ensured that the data set was kept in analyzable sizes and the data processing process was manageable [17, 18]. However, the dataset is limited to a one-month time frame and does not include demographic or geographic metadata due to API constraints. Therefore, it may not fully represent the broader population. Future studies should incorporate longer temporal coverage and user-level metadata to improve representativeness.

In this process, a total of 2,046 Turkish tweets and 18,000 English tweets were collected. During the preprocessing phase, the dataset was systematically refined by removing duplicate entries, noise, incomplete and inconsistent records, as well as irrelevant content, to enhance data quality, reduce potential bias, and ensure the robustness and reliability of subsequent analyses. These procedures are aimed at increasing the accuracy of the analysis process and to obtain meaningful results. After the cleaning process, the number of tweets suitable for analysis was determined to be 1104 for Turkish and 6449 for English. The data cleaning process involved removing duplicate tweets, non-relevant content (such as advertisements), tweets containing only URLs or hashtags, and posts with insufficient textual information. Additionally, tweets with excessive noise, including random characters and non-linguistic symbols, were excluded.

To minimize potential bias, the same cleaning criteria were applied consistently across both datasets. Furthermore, sentiment distributions were checked before and after preprocessing, and no significant distortion in class proportions was observed, suggesting that the filtering process did not systematically bias the results.

Table I presents examples from the dataset, including multilingual tweets addressing climate change issues.

TABLE I. SAMPLE DATASET

Tweet	Language
Whatever the cause, climate change seems to be an area where countries will work together and struggle to maintain a humane life through joint decisions.	Tr
We must stop the development of agricultural land. Global warming and climate change will make it extremely difficult for people to access healthy food and water.	Tr
In recent years, dam occupancy rates have started to decrease due to reasons such as climate change, drought and increase in water use.	Tr
Serious policies should be developed on climate change and drought. Yes, to green policies for a more loveable world...	Tr
Our planet is the only place where humanity lives and therefore its protection is of utmost importance. However, human activities have resulted in several environmental problems such as overconsumption of natural resources, environmental pollution, climate change and habitat loss.	Tr
The climate crisis forces us to reconsider our values.	En

Humanity must fight climate change much harder, or losses and damage, and suffering, will continue to mount.	En
There's no solution to climate change - and no one can do it alone	En
Climate change is the threat that will unite us all.	En
Scientists revive viruses frozen thousands of years ago in the Arctic to predict the dangers that could be unleashed by melting because of climate change	En

B. *Text Mining*

Text mining is a big data analysis method that generates meaningful information from text data. Text mining is a data mining study whose data source is text, and it enables the transformation of unstructured data into structured data. Text mining is the process of extracting confidential information from the data in the text that does not have a clear format and formatting the disorganized data [19].

The first thing to do with the data in the analysis process is to pre-process the data. In a data mining process, previous studies have shown that the preprocessing process covers up to 60% of the necessary work, and preprocessing the data contributes 75% to 90% to the success of the data mining project. Non-textual characters such as !, ?, /, +, and emoji in the documents are removed from the data set. The stop words that occur in large numbers in the documents are extracted based on the previously prepared stop word list. Since the use of ineffective words during indexing reduces the size of the index files, it increases memory usage and query processing efficiency [20].

N-gram is a method used to search and compare data or to determine the number of repetitions. Term frequency is used to calculate the term weights in a document. Reverse term frequency is used to find the number of occurrences of the word in more than one document and tries to understand whether this word is a term or a conjunction, etc. (stop-words). It is found by taking the absolute value of the logarithm of the number of documents in which the term occurs divided by the number of documents [21].

These feature extraction methods contribute to model performance by enhancing semantic representation of textual data. However, alternative feature representations such as contextual embeddings (e.g., BERT) were not evaluated, which constitutes a limitation of the study.

C. *Sentiment Analysis*

To train the supervised LSTM model, sentiment labels were generated using a semi-automated approach. Initially, tweets were labelled using lexicon-based sentiment resources and existing benchmark datasets. Subsequently, a subset of the data was manually reviewed to ensure consistency and correctness of the assigned labels. This approach improves the reliability of the training data and reduces potential bias arising from fully automated labelling. Sentiment analysis is the field of study that analyses people's opinions, emotions, evaluations, and attitudes through written language. Users' thoughts about a particular event are extracted through text analysis, and then emotions are predicted based on the analysis. Due to the increasing data flow, sentiment analysis has become one of the most active topics of natural language processing and text mining [22].

Sentiment analysis studies are questioned and analyzed whether the texts have positive, negative, or neutral content. According to the results of this analysis, the attitude of individuals or a certain group on the subject related to the study is determined [23].

The methods used in detecting sentiment polarity fall under two main categories. These categories are machine learning-based methods and word-based methods. In machine learning-based methods, the system is trained with pre-labelled training data, and emotion classification is performed with the trained system. Any of the machine learning classification algorithms can be used. Under this category, methods based on syntactic analysis of sentences using natural language processing methods and tools are used. There is no need for labelled training data as in supervised machine learning methods. With natural language processing tools and methods, sentences are analyzed, and semantic inferences are made by detecting emotion terms in sentences. A dictionary of emotion terms is often used to detect emotion expressions in sentences [24, 25].

D. *Machine Learning*

Machine learning is an important subfield of artificial intelligence. Its main purpose is to use computational methods to extract knowledge from data. Machine learning has a wide range of applications, including handwriting and speech recognition, robotics and computer games, natural language processing, brain-machine interfaces, etc. [26].

The classification method, one of the basic methods of data mining, is based on the learning algorithm [27]. Classification algorithms divide the data into certain groups (classes) according to their common characteristics by targeting the desired information. Classification algorithms are run in two stages. Firstly, the classification model is created by analyzing the data set determined as training data. In the second stage, the obtained classification model is applied to a new dataset, and the presence of the determined classes in the data is investigated [28].

Deep learning is a subset of machine learning that enables models to learn hierarchical representations from data through multiple layers of neural networks [29]. Unlike traditional machine learning techniques that rely on handcrafted features, deep learning models automatically extract relevant features from raw text data. In the context of sentiment analysis, deep learning provides a powerful framework for capturing complex language patterns and contextual meanings in social media posts [30].

One of the primary advantages of deep learning for text analysis is its ability to process sequential data effectively. Social media posts, such as tweets, are inherently sequential and often contain nuanced expressions, abbreviations, and informal language structures [31]. Deep learning models, particularly those based on recurrent neural networks (RNNs) and transformer architectures, have demonstrated superior performance in sentiment classification tasks compared to traditional machine learning methods [32].

Long Short-Term Memory (LSTM) networks stand out as a specialized type of recurrent neural network (RNN) particularly well-suited for processing sequential data [33]. Unlike traditional RNNs, which struggle with long-range dependencies due to the vanishing gradient problem, LSTM networks incorporate memory cells and gating mechanisms to retain relevant information over extended sequences. This capability makes LSTM an ideal choice for analyzing textual data, where contextual relationships between words play a crucial role in sentiment classification.

The architecture of an LSTM network consists of three key gates [34]:

- Forget Gate: Determines which information from the previous time step should be discarded.
- Input Gate: Controls how much new information should be added to the memory cell.
- Output Gate: Regulates the amount of stored information that contributes to the current output.

For this study, the LSTM-based sentiment analysis model was trained on a dataset of Turkish and English tweets related to climate change. The preprocessing pipeline involved tokenization, stop-word removal, and word embedding using pre-trained word vectors. The model was structured with multiple LSTM layers, dropout regularization to prevent overfitting, and a SoftMax activation function in the output layer to classify tweets into three sentiment categories.

The training process involved optimizing the categorical cross-entropy loss function using the Adam optimizer. The dataset was split into training and validation sets to evaluate model performance [35]. Additionally, performance metrics such as accuracy, precision, recall, and F1-score were computed to assess the effectiveness of the LSTM model in sentiment classification.

The deep learning models were implemented using TensorFlow and Keras libraries. The input text data was vectorized using word embeddings such as Word2Vec and GloVe to enhance semantic understanding. Hyperparameters, including batch size, learning rate, and the number of LSTM units, were fine-tuned through grid search and cross-validation [36].

To ensure robustness, the model was trained on a diverse dataset comprising tweets in both Turkish and English. The bilingual nature of the dataset posed challenges in terms of language-specific features; however, LSTM's ability to capture sequential dependencies allowed for effective sentiment classification across languages. Data augmentation techniques were also employed to balance class distributions and improve generalization [37].

By leveraging deep learning techniques, particularly LSTM networks, this study effectively classified climate change-related tweets based on sentiment polarity. The proposed approach outperformed traditional machine learning methods by capturing complex linguistic patterns and long-term dependencies in textual data. The insights derived from this analysis provide valuable contributions to understanding public perceptions of climate change and developing targeted environmental communication strategies. It should be noted that this study does not employ explicit cross-lingual embedding alignment or normalization techniques. Instead, Turkish and English datasets are analyzed independently to preserve language-specific characteristics. While this approach limits direct semantic comparability, it enables the identification of natural linguistic differences between languages.

E. Topic Modelling

Topic modelling (TM) is one of the subfields of text mining. Its main goal is to reveal hidden or open topics in documents. It has an important place among the subfields of text mining, especially with its increasing importance and studies carried out in recent years. KM is a method in which subject classification is made unsupervised in a document [38].

Topic modelling methods are based on the principle that topics with probability distributions on words come together randomly and form documents. Latent Dirichlet Allocation (LDA) is a generative graphical model used to model discrete data such as documents and to reveal the topics that make up the document. "Latent" here refers to finding the meaning of the document by discovering the hidden topics that make up the document. What is meant by generative model is the generation of the words in the document by a simple probabilistic process within the framework of hidden (random) variables, i.e., the creation of the document. GDA, which is a completely unsupervised method, does not require any prior knowledge and works based on a bag-of-words approach. While the placement of words in the document is ignored, the co-occurrence of words is used in this method [39].

Each document consists of a random mixture of topics, and each of the words that make up the document is chosen from one of the topics. The topics are also probability distributed from words in a fixed dictionary.

The basic components of the LDA model are as follows [39]:

- Number of topics (K): This specifies the number of topics that the model will predict.
- Document topics (θ_d): Each document is represented as a mixture of K topics. Here, θ_d is the topic distribution of the d-th document.
- Topic word distributions (ϕ_k): Each topic represents a distribution of words. That is, ϕ_k is the word distribution of the k-th topic.
- Word vector (w_d, n): The words that make up each document. Each word is generated by a topic.

The mathematical structure of the LDA model can be expressed as follows:

$$p(w, z\, I\alpha, \beta) = \prod_{d=1}^{D}\left(\prod_{n=1}^{N_d} p\left(\left(w_{d,n}|z_{d,n}, \beta\right)p\left(z_{d,n}|\theta_d\right)\right)\right)p(\theta_d|\alpha)p(\beta|\eta) \quad (1)$$

where, w_d, n is the n-th word in the d-th document, and z_d, n is the predicted topic label for this word. θ_d represents the topic distribution of the d-th document. β and α are hyperparameters of Dirichlet distributions.

IV. RESULTS

Fig. 1 shows the block diagram of the study, illustrating the sequential stages of data collection, preprocessing, and analysis.

Fig. 1. Block diagram.

In total, 2046 tweets in Turkish and 18,000 tweets in English were initially collected from Twitter using relevant climate change-related keywords. The dataset underwent a rigorous preprocessing phase, including text cleaning, tokenization, and stop-word removal, to enhance the quality and reliability of the textual data. Following this process, tweets deemed unsuitable for analysis due to excessive noise, irrelevance, or duplication were excluded. As a result, the final dataset comprised 1104 Turkish tweets and 6449 English tweets that met the criteria for further sentiment and topic analysis.

The refined dataset was then subjected to deep learning-based sentiment classification and topic modelling to uncover prevalent themes and emotional tendencies in climate change discussions. The subsequent sections present the findings of the sentiment analysis, topic modelling results, and a comparative evaluation of Turkish and English tweets regarding climate change perceptions.

To classify the tweets according to their sentiment using machine learning methods, the data set was divided into two as training and test. Seventy percent of the data set is divided into training and 30 percent test data. At this stage, training and test sets were randomly generated to ensure data diversity. In all the algorithms used for sentiment analysis, 70% of the data set allocated for training is used while creating the model. After the model-building process, 30% of the test data is used to predict the model, and the performance of the model was evaluated.

A. Sentiment Analysis

Deep learning, a subset of machine learning, is used for sentiment analysis. Classification is the process of assigning a data set to the most appropriate one of the different and predetermined training data categories. The process of classifying the tweets according to their sentiments was carried out according to the climate change comments in the tweets. The aim of this study is to classify tweets about climate change with three emotions: positive, negative, and neutral.

Precision, recall, accuracy, and F1 metrics are used to evaluate the success of the model. TP (True Positives) refers to true positive predictions, FP (False Positives) refers to false positive predictions, FN (False Negatives) refers to false negative predictions, and TN (True Negatives) refers to true negative predictions [15]:

Precision indicates how accurate the result is in the model created.

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

Recall shows the ability to find the correct examples in the created model.

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

Accuracy indicates the success of the model in predicting the target classes.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

F1 Score is the harmonic mean of precision and recall.

$$F1 = \frac{2 * Recall * Precision}{Precision + Recall} \tag{5}$$

Table II shows the deep learning performance evaluation.

TABLE II. PERFORMANCE MEASUREMENT

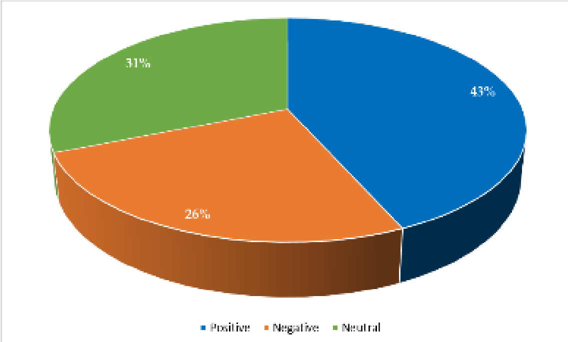
Metrics	English Tweet	Turkish Tweet
Precision	0.912	0.935
Recall	0.932	0.946
Accuracy	0.903	0.924
F Score (F1)	0.922	0.940

The deep learning model performed well on both English and Turkish tweets, with the metrics for Turkish tweets being generally better. Precision and recall values show that the model correctly captures positive classifications and misclassifications are low in both languages. For Turkish tweets, precision 0.935, recall 0.946 and F1 score 0.940 are higher than English tweets (0.912, 0.932 and 0.922, respectively). In terms of accuracy, Turkish tweets (92.4%) performed better than English tweets (90.3%). These results suggest that the language features or data structure of Turkish tweets may be better learnt by the model and that the model works more effectively with Turkish data. However, no statistical significance testing or confidence interval estimation was conducted. Future research should include hypothesis testing and variance analysis to assess the robustness of the observed performance differences. Table III shows the sentiment distribution of the tweets.

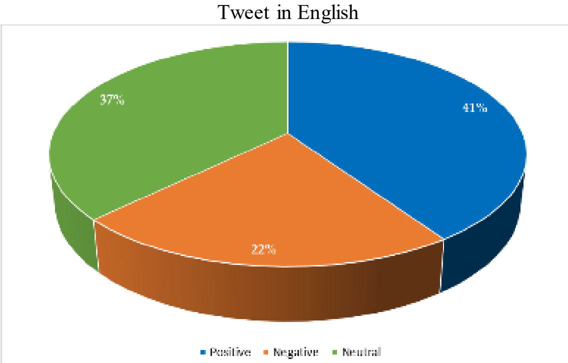
TABLE III. SENTIMENT STATISTICS

Sentiment Analysis	Tweet Number	
	EN	TR
Positive	2795	447
Negative	1646	246
Neutral	2007	411

The sentiment analysis results in Table III show that there are similar distributions between English and Turkish tweets. In both English and Turkish tweets, positive sentiment has a higher rate than other categories. Fig 2 shows the distribution of sentiment in terms of percentage.



Tweet in English



Tweet in Turkish

Fig. 2. Sentiment mood distribution.

In Fig. 2, 43.3% of the tweets in English were positive, 25.5% were negative, and 31.2% were neutral, while these rates were 40.5% positive, 22.3% negative, and 37.2% neutral in Turkish tweets, respectively. These distributions show that a generally positive or neutral perspective on climate change is dominant in both languages. However, the slightly higher rate of neutral sentiments in Turkish tweets compared to English tweets may suggest that Turkish-speaking users approach the issue with a more neutral or observant approach. In English tweets, the difference between positive and negative sentiments is more pronounced than in Turkish, which may indicate that English-speaking users may tend to have a stronger emotional response. However, this interpretation should be treated with caution, as the dataset is limited to Twitter users who shared posts using specific climate change-related keywords within a defined time frame.

B. Text Mining

The n-gram and term frequency and inverse term frequency methods are used to determine the core attributes. Word frequencies show the number of times words are used in the data set. Word frequencies are shown using a collection of matrix terms from the most frequently used words to the least frequently used words in the dataset. Term Frequency (TF) is the frequency of repetition of a term within a document. Since each document is different in length, a document is likely to appear much more often in longer documents than in shorter ones. Inverse Document Frequency (IDF) is the importance of a term in the entire document collection (D). According to IDF, the importance of a term is directly proportional to the frequency with which the term is used in the document and inversely proportional to the number of times the term is used in the entire document pool. Low-frequency terms have a high IDF score; high-frequency terms have a low IDF score. The term frequency-inverse text frequency (TF-IDF) value takes a high value if the term occurs a lot in a small number of documents. If the term is used in all documents, the TF-IDF value takes its lowest value [40].

Table IV shows the most frequent words extracted from English tweets related to the research topic.

TABLE IV. FREQUENCY LIST OF ENGLISH TWEETS

Positive	Negative	Neutral
(“China”, 153)	(“China”, 98)	(“China”, 185)
(“World”, 116)	(“Global”, 77)	(“Energy”, 152)
(“Energy”, 91)	(“Energy”, 49)	(“Global”, 50)
(“Carbon”, 77)	(“Money”, 39)	(“Money”, 48)
(“Natural”, 70)	(“Fossil”, 39)	(“Help”, 44)
(“Weather”, 61)	(“Government”, 37)	(“Warming”, 28)
(“Power”, 57)	(“Weather”, 36)	(“Weather”, 28)
(“Environmental”, 56)	(“Food”, 32)	(“Government”, 28)
(“Fossil”, 53)	(“Water”, 32)	(“Control”, 24)
(“Government”, 50)	(“Warming”, 32)	(“Carbon”, 22)

The analysis of positive, negative, and neutral words in English tweets in Table IV shows that discussions on climate change have a multidimensional structure. Among the positive words, terms such as “China” and “energy” point to promising developments such as renewable energy investments and carbon-neutral targets, while expressions such as “world” and “natural” reflect a positive view of global cooperation and environmentally friendly approaches. On the other hand, in the negative category, words such as “China,” “fossil,” and “government” represent criticism of carbon emissions, dependence on fossil fuels, and political inefficiencies. Neutral words, on the other hand, usually consist of terms used in neutral contexts such as “energy”, “global,” and “money,” and are used to describe news, policies, or general situation assessments related to climate change. These data reveal that, as stated in the literature [6], public perception of climate change is balanced between both positive and negative emotions and neutral information.

Table V shows the most frequent words extracted from Turkish tweets related to the research topic.

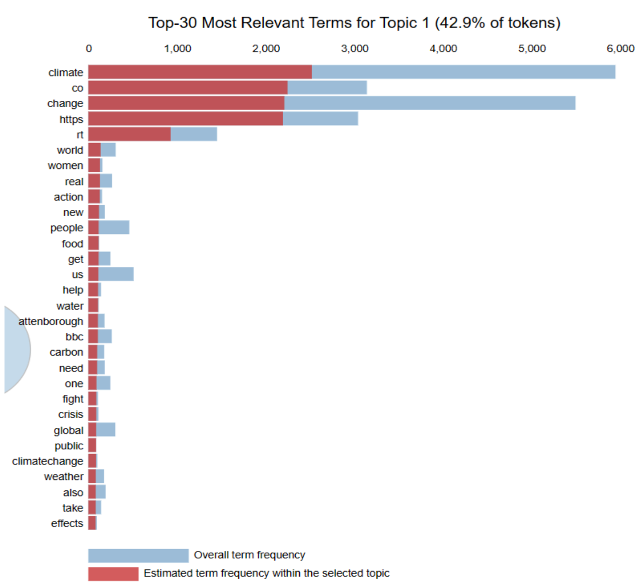
TABLE V. FREQUENCY LIST OF TURKISH TWEETS

Positive	Negative	Neutral
(“Environment,”, 106)	(“Global”, 60)	(“Urbanization”, 71)
(“Science”, 96)	(“Food”, 22)	(“Environment,”, 70)
(“Minister”, 89)	(“Companies”, 21)	(“Minister”, 49)
(“Urbanization”, 74)	(“Earthquake”, 17)	(“Presidency”, 32)
(“Global”, 40)	(“Environment,”, 15)	(“Education”, 18)
(“Warming”, 30)	(“Damage”, 13)	(“Disaster”, 31)
(“Water”, 24)	(“Warming”, 12)	(“Water”, 15)
(“Earth”, 24)	(“Water”, 12)	(“Management”, 12)
(“Economic”, 17)	(“Population”, 11)	(“Municipality”, 10)
(“Temperatures”, 14)	(“Carbon”, 9)	(“Cost”, 29)

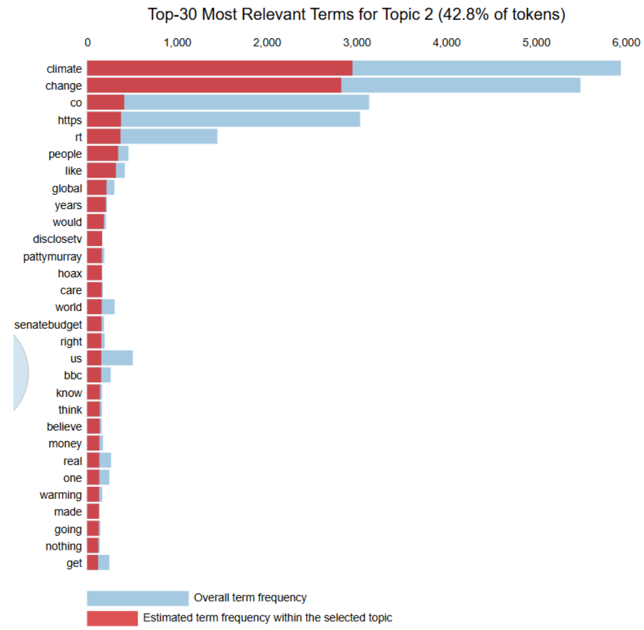
In Table V, the most common positive, negative, and neutral words in Turkish tweets in the context of climate change show that the perception of this issue has a multi-faceted structure. "Environment" and "science", which stand out among the positive words, reflect promising approaches to environmental and scientific issues, while terms such as "urbanization" and "minister" show that urbanization and initiatives taken at the leadership level are discussed in a positive framework. In contrast, words in the negative category, such as "global" and "food," represent concerns about global challenges and the food crisis. Expressions such as "earthquake" and "damage" indicate negative perceptions about the disasters triggered by climate change and their consequences. Neutral words such as "presidency," "management," and "disaster" are terms generally used in political, administrative, and crisis management discussions. The frequency of use of these words reveals that social and managerial sensitization about climate change has increased at both local and global levels. As stated in the literature [6], messages about climate change in the media and social media are mostly presented as a balance of both positive (solution-oriented approaches) and negative (worrisome effects) content. In this context, the discussions in Turkish tweets show that local problems (water resources, natural disasters) and solution proposals (scientific approaches, leadership) form a blended narrative.

C. Topic Modelling

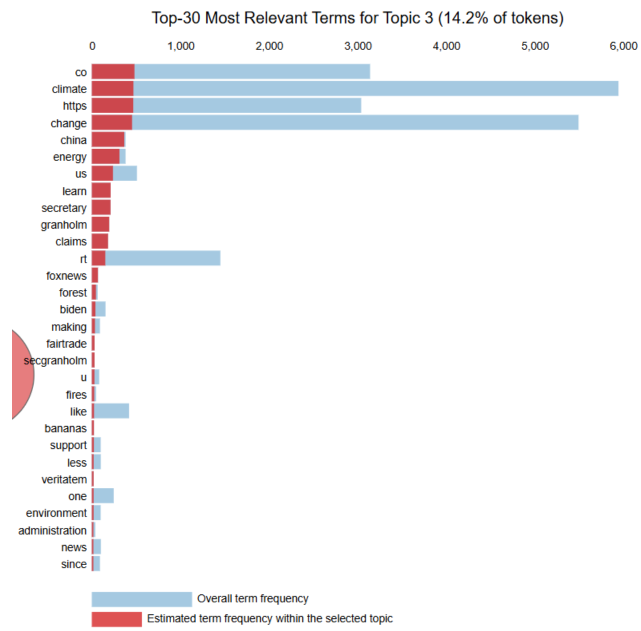
The algorithms proposed for topic modelling are statistical methods that aim to reach a conclusion by analyzing the words that make up the document. They do not need any labelling step while discovering the connection between the topics and the changes they show over time. While the placement of words in the document is ignored, the co-occurrence of words is used in this method. In this study, topic modelling is performed on 3 topics using Latent Dirichlet Allocation (LDA). The number of models was decided by trial and error, and the most appropriate number of topics was determined as 3. Fig. 3 shows the topic modelling of English tweets, and Fig. 4 shows the topic modelling of Turkish tweets.



Topic 1



Topic 2



Topic 3

Fig. 3. LDA terms (English tweets).

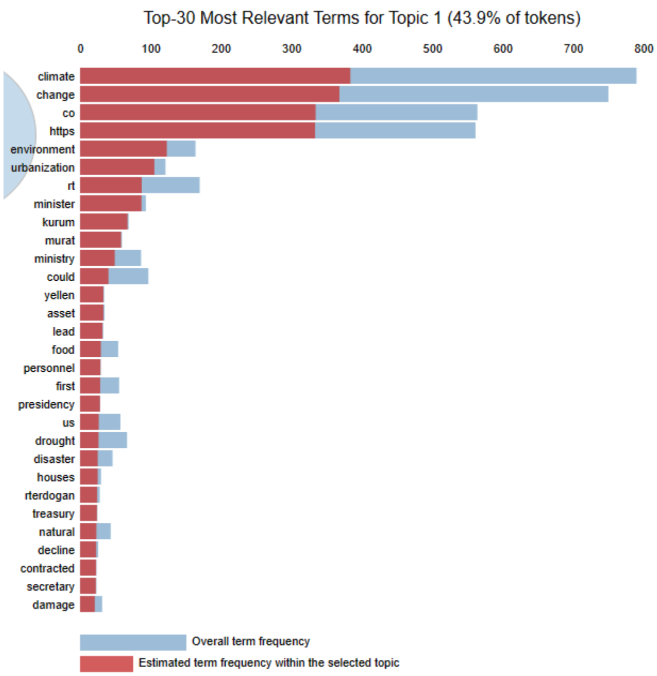
The three topic distributions of the English tweets in Fig. 3, which resulted from the topic modelling, are as follows:

Topic 1: Global Awareness and Calls to Action

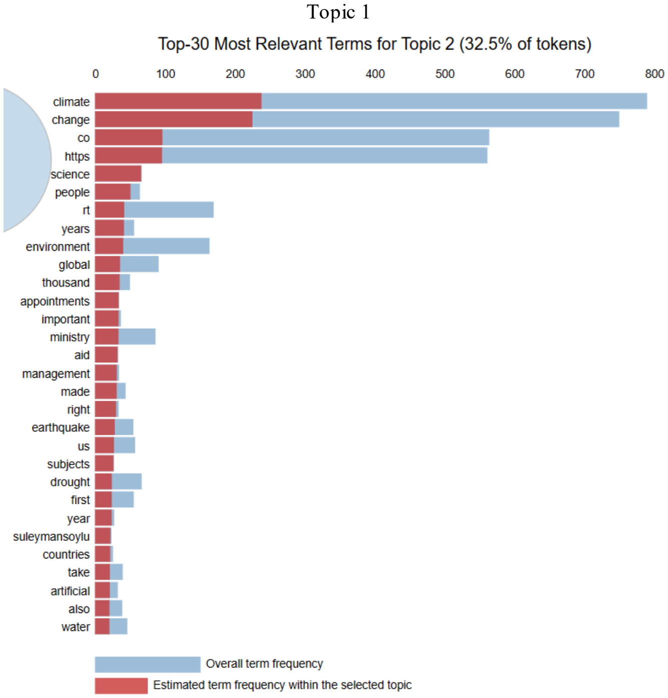
Topic 1 centers around awareness and calls for action on climate change. While the highest frequency of use of the term's "climate", and "change" constitutes the main theme of this topic, terms such as world, women, real, action, help, and public indicate that the global impacts of climate change are emphasized. Under this heading, awareness-raising and calls for action on the environmental crisis and broader societal issues come to the fore.

Topic 2: Climate Change Debates and Polarized Views

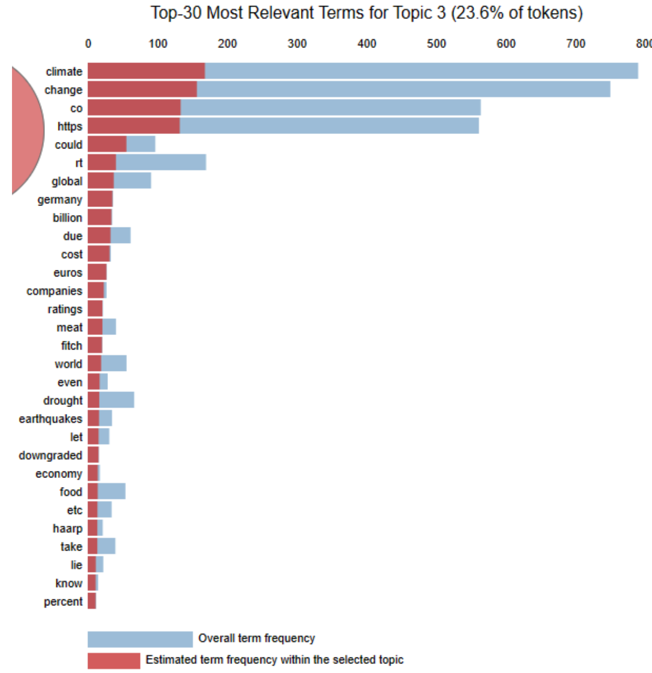
Topic 2 reflects the reality of climate change and the controversial discourses around it. While “climate”, and “change” are again important keywords, terms such as “hoax”, “believe”, think”, “care”, “right”, and “senate budget” points to the differences of opinion between those who believe in climate change and those who question it. In addition, terms such as money and economics indicate that the economic dimensions of climate policies also occupy an important place in the debates.



Topic 1



Topic 2



Topic 3

Fig. 4. LDA terms (Turkish tweets).

Topic 3: Political Discourses and Energy Policies

Topic 3 shows that climate change is addressed in the context of political actors, media, and energy policies. The prominence of terms such as “China”, “energy”, “secretary”, “Granholm”, “Fox News”, and “administration” shows that climate change discourses are shaped through political figures, media organizations, and energy policies. Under this heading, discussions on energy policies and the role of specific countries or political actors come to the fore.

The three topic distributions of the Turkish tweets in Fig. 4, which resulted from the topic modelling, are as follows:

Topic 1: Political and Institutional Discourses and Climate Change

This topic reflects political, bureaucratic, and institutional discourses on climate change. Terms such as “minister”, and “institution” indicate the intensification of public debates over the statements or policies of government officials and public institutions on this issue. Expressions such as “environment”, “urbanization”, and “natural disaster” reveal that environmental impacts, sustainable urbanization debates, and natural disasters are addressed in a political context.

Topic 2: Scientific Perspectives and Climate Crisis in a Global Framework

This heading includes scientific and global discourses on the climate crisis. Terms such as “science”, “environment”, and “global” indicate that scientific studies and international perspectives are frequently emphasized in tweets. In addition, terms such as “earthquake”, “drought”, and “water” emphasize natural disasters and resource problems caused by the climate crisis. In this context, the issue can be interpreted as an area where climate change is discussed through scientific findings and global impacts.

Topic 3: Economic Impacts and Social Reactions

The third issue shows that climate change is related to economic impacts and societal perceptions. Expressions such as “economy”, “cost”, “companies”, and “billion” reveal that economic risks and costs are frequently mentioned. Financial terms such as “downgraded” and “ratings” reflect the impacts of climate change on corporate and economic performance. In addition, terms such as “global” and “drought” indicate that this economic impact has broad repercussions at the global level.

V. Discussion

This study’s analysis of social media perceptions of climate change makes important contributions to the literature by comparatively evaluating the contextual and emotional aspects of Turkish and English tweets. The findings obtained through deep learning-based sentiment analysis and topical modelling methods show that individuals” perceptions of climate change are influenced not only by linguistic but also cultural contexts. Accordingly, the study provides important results to better understand the nature of discussions on climate change in social media and to provide a roadmap for the development of environmental communication strategies.

The result of the study shows that emotional reactions were stronger in English tweets, while a more neutral attitude was more prominent in Turkish tweets. This finding coincides with the literature showing the relationship between cultural values, media usage habits and environmental sensitivities. [28] stated that social media plays an important role in public debates on climate change and that culture-specific perception differences are reflected in social media content.

While the higher proportion of neutral expressions in Turkish content indicates that users adopt an observant approach, the emotional intensity in English tweets reveals that the Western public adopts a more activist stance on environmental issues. This was also confirmed [37] in a cross-cultural study covering 110 countries; it was stated that the public perceived climate change as a higher risk and reactions to it were higher in more affected geographies.

Social media stands out as a space where individuals can make their voices heard on environmental issues, access information and play a role in shaping policies. In this context, the concept of “Fifth Power” articulated [10] emphasizes the capacity of individuals to generate knowledge and social intervention through digital platforms. In the tweets analyzed in this study, especially in English content, frequently used expressions (e.g., “action”, “real”, “help”) can be considered as an indicator of digital activism.

In particular, the “Greta Effect”, which symbolizes the impact of climate activist Greta Thunberg, has shown the power of online campaigns to mobilize decision-makers [20]. The findings obtained in the study reveal that similarly, large masses can be reached, and awareness can be raised through social media.

The use of machine learning and deep learning methods provides significant advantages in analyzing social media data. The LSTM model used in this study classified both Turkish and English tweets with high accuracy rates. This supports that deep learning architectures are more successful than traditional methods in text classification, as stated [35].

Higher F1 and accuracy rates for Turkish tweets indicate that the model may have learnt the language features better. This finding is in line with the studies conducted [3] on Turkish content classification. Although the Turkish dataset is smaller than the English dataset, the higher performance may be explained by differences in data characteristics. The Turkish dataset appears to be more homogeneous and less noisy, which can facilitate pattern learning for the model. In contrast, the larger English dataset likely includes more linguistic variability, sarcasm, and contextual ambiguity, which may reduce classification performance. Additionally, differences in class distribution may also contribute to this outcome.

The results of the topical modelling show that discussions on climate change are shaped around political, economic and scientific perspectives. While energy policies, media discourses and the role of political figures are prominent in English tweets, public institutions” statements and scientific discourses are more dominant in Turkish tweets. These differences reveal that environmental communication strategies should be customized according to language and context.

Especially words such as “hoax”, “senate”, and “budget” in English tweets reflect the polarization and political debates on climate change. [16] stated that such polarization weakens the consensus on climate change on social media and facilitates the spread of misinformation. These findings are limited to the analyzed dataset and should not be generalized to broader populations. The results reflect only the behavior of Twitter users within the selected time frame and keyword scope.

VI. Conclusions

The study provides an important perspective on this issue by analyzing social perceptions and emotional tendencies regarding climate change in Turkish and English tweets. Using artificial intelligence-based text mining and sentiment analysis methods, the study shows that discussions about climate change on social media platforms have a multidimensional structure. While the difference between positive (43.3%) and negative (25.5%) sentiments is more pronounced in English tweets, neutral sentiments (37.2%) have a higher rate in Turkish tweets. This suggests that English-speaking users tend to have stronger emotional reactions to climate change, while Turkish-speaking users tend to adopt a more observant or neutral approach. In the positive statements, words such as energy, China, and world point to promising developments such as renewable energy investments and global cooperation, while the words fossil, economy, and government in the negative category reflect criticisms such as fossil fuel dependency and political inefficiencies. These data support the findings in the literature that social perception of climate change is based on a balance of positive and negative emotions [37].

Nevertheless, these findings are specific to the sampled social media population and should not be generalized to broader linguistic communities. The results reflect only the behavior of users active on Twitter within the selected dataset and keyword scope.

In this study, deep learning-based models have been successfully used for sentiment analysis related to climate change. Natural language processing (NLP) techniques and deep learning methods were effectively applied to accurately classify the emotional content of Twitter data. The deep learning model LSTM (Long Short-Term Memory) offered higher accuracy rates than traditional methods in understanding the context of texts and accurately capturing emotional tones. In this way, it was possible to analyses users" feelings and thoughts about climate change in more depth and to compare social media data in different languages (Turkish and English). The high accuracy of deep learning methods in such analyses reveals the complex emotional structure of social media posts more clearly.

The importance of the article stems from the fact that it presents important implications at both the local and global levels by analyzing the perceptions and attitudes of individuals on climate change and climate change issues through social media data. Climate change is seen as a critical issue not only for environmental sustainability but also for the future of economic and social policies. Therefore, sentiment analysis of the content shared on social media about climate change provides valuable information to understand the public's concerns, hopes, and demands on this issue. The comparison of Turkish and English tweets has an important originality in terms of showing how perceptions towards climate change are shaped in different cultural contexts.

However, the study also has some limitations. Firstly, the fact that only Twitter data is used makes it difficult to generalize the results to other social media platforms. For example, a more comprehensive analysis can be made with content obtained from platforms such as Instagram, Facebook, or TikTok. Furthermore, the fact that only text-based analyses were conducted, and visual content or hashtag trends were not examined limited the possibility of making an evaluation covering all dimensions of social media posts. Another limitation is that the analysis is based on keyword frequencies regardless of the context of the texts. This may prevent a full understanding of users" real intentions and emotional states.

In future studies, comparative analyses of different social media platforms can be included, and the accuracy of sentiment analysis can be improved by using more advanced deep learning methods. In addition, the attitudes of different demographic groups (age, gender, geographical region) towards climate change can be analyzed, and more detailed inferences can be made on how local and global policies are perceived. By extending the study, more inclusive and effective strategies can be developed, especially for policymakers and stakeholders that drive climate change investments. Such analyses will play a critical role in both formulating policies in line with the demands of the public and raising awareness on combating climate change. Future studies may extend this work by incorporating multimodal data (e.g., images and videos) and developing real-time monitoring systems to track public perception of climate change.

REFERENCES

[1] IPCC, Climate Change 2023: Synthesis Report (Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change), 2023.

[2] S. Lewandowsky, J. Cook, N. Fay, and G. E. Gignac, "Science by social media: Attitudes towards climate change are mediated by perceived social consensus," Memory & Cognition, vol. 47, pp. 1445–1456, 2019.

[3] S. R. Weart, The Discovery of Global Warming, Harvard University Press, Cambridge, MA, 2005.

[4] S. C. Moser, "Communicating climate change: history, challenges, process and future directions," Wiley Interdisciplinary Reviews: Climate Change, vol. 1, no. 1, pp. 31–53, 2010.

[5] W. Pearce, S. Niederer, S. M. Özkula, and N. Sánchez Querubín, "The social media life of climate change: Platforms, publics, and future imaginaries," Wiley Interdisciplinary Reviews: Climate Change, vol. 10, no. 2, 2019.

[6] W. H. Dutton, The Fifth Estate: The Power Shift of the Digital Age, Oxford University Press, 2023.

[7] J. Bogert, J. Buczny, J. Harvey, and J. Ellers, "The effect of trust in science and media use on public belief in anthropogenic climate change: A meta-analysis," Environmental Communication, 2023.

[8] J. Thaker, "Cross-country analysis of climate change perception and protest participation," Environmental Communication, vol. 18, no. 3, pp. 233–249, 2024.

[9] N. G. Mede and R. Schroeder, "The "Greta effect" on social media: A systematic review," Environmental Communication, 2024.

[10] L. Dellmuth and K. Shyrokykh, "Climate change on Twitter: Implications for climate governance research," Wiley Interdisciplinary Reviews: Climate Change, vol. 14, no. 6, e848, 2023.

[11] J. Burger, M. Gochfeld, T. Pittfield, and C. Jeitner, "Perceptions of climate change and consequences," Energy and Power Engineering, vol. 8, no. 5, p. 250, 2016.

[12] L. Casiraghi, E. Kim, and N. Hara, "Tweeting on thin ice: Scientists in dialogic climate change communication," First Monday, 2024.

[13] J. Pueyo-Ros and E. Garau, "Characterizing attitudes towards climate change via sentiment analysis," Journal of Integrative Environmental Sciences, vol. 20, no. 1, 2023.

[14] L. Whitmarsh, "Risk perception and behavioural response to climate change," Journal of Risk Research, vol. 11, no. 3, pp. 351–374, 2008.

[15] A. Rousseau, "Climate-related social media engagement," Communication Research, 2023.

[16] X Platform, "X platform," 2024. [Online]. Available: https://x.com

[17] K. Makice, Twitter API: Up and Running, O"Reilly Media, 2009.

[18] M. Trupthi, S. Pabboju, and G. Narasimha, "Sentiment analysis on Twitter using streaming API," IEEE International Conference, 2017.

[19] M. Pejić Bach, Ž. Krstić, S. Seljan, and L. Turulja, "Text mining for big data analysis," Sustainability, vol. 11, no. 5, p. 1277, 2019.

[20] S. Vijayarani, M. J. Ilamathi, and M. Nithya, "Preprocessing techniques for text mining," International Journal of Computer Science & Communication Networks, vol. 5, no. 1, pp. 7–16, 2015.

[21] M. Schonlau, N. Guenther, and I. Sucholutsky, "Text mining with n-gram variables," The Stata Journal, vol. 17, no. 4, pp. 866–881, 2017.

[22] B. Liu, Sentiment Analysis and Opinion Mining, Springer, 2022.

[23] E. Martínez-Cámara, M. T. Martín-Valdivia, L. A. Ureña-López, and A. R. Montejo-Ráez, "Sentiment analysis in Twitter," Natural Language Engineering, vol. 20, no. 1, pp. 1–28, 2014.

[24] Y. I. Alzoubi, A. E. Topcu, and A. E. Erkaya, "Machine learning-based text classification," Applied Sciences, vol. 13, no. 16, p. 9428, 2023.

[25] A. Palanivinayagam, C. Z. El-Bayeh, and R. Damaševičius, "Machine-learning-based text classification: A systematic review," Algorithms, vol. 16, no. 5, p. 236, 2023.

[26] I. H. Sarker, “Machine learning: Algorithms and applications,” SN Computer Science, vol. 2, no. 3, p. 160, 2021.

[27] S. S. Nikam, “Classification techniques in data mining,” Oriental Journal of Computer Science and Technology, vol. 8, no. 1, pp. 13–19, 2015.

[28] J. Tang, S. Alelyani, and H. Liu, “Feature selection for classification,” in Data Classification: Algorithms and Applications, 2014.

[29] R. B. Palm, “Prediction as a candidate for learning deep hierarchical models,” Technical University of Denmark, 2012.

[30] A. Yadav and D. K. Vishwakarma, “Sentiment analysis using deep learning architectures,” Artificial Intelligence Review, vol. 53, no. 6, pp. 4335–4385, 2020.

[31] S. F. Ahmed et al., “Deep learning modelling techniques,” Artificial Intelligence Review, vol. 56, no. 11, pp. 13521–13617, 2023.

[32] K. L. Tan, C. P. Lee, K. S. M. Anbananthen, and K. M. Lim, “RoBERTa-LSTM hybrid model,” IEEE Access, vol. 10, pp. 21517–21525, 2022.

[33] N. M. Rezk, M. Purnaprajna, T. Nordström, and Z. Ul-Abdin, “Recurrent neural networks,” IEEE Access, vol. 8, pp. 57967–57996, 2020.

[34] J. Khan, N. Ahmad, S. Khalid, F. Ali, and Y. Lee, “Hybrid DNN for sentiment classification,” IEEE Access, vol. 11, pp. 28162–28179, 2023.

[35] Y. Ho and S. Wookey, “Cross-entropy loss function,” IEEE Access, vol. 8, pp. 4806–4813, 2019.

[36] V. Yesilevskyi and M. Kyt, “AI education trends,” International Journal of Emerging Technologies in Learning, vol. 19, no. 4, pp. 86–96, 2024.

[37] K. Munir, H. Zhao, and Z. Li, “Semantic role labeling,” IEEE/ACM Transactions, vol. 29, pp. 782–791, 2021.

[38] M. Mihalovic, “Classification models performance,” Economics & Sociology, vol. 9, no. 4, p. 101, 2016.

[39] H. Jelodar et al., “Latent Dirichlet allocation and topic modeling,” Multimedia Tools and Applications, vol. 78, pp. 15169–15211, 2019.

[40] H. Christian, M. P. Agus, and D. Suhartono, “TF-IDF for text mining,” ComTech, vol. 7, no. 4, pp. 285–294, 2016.