A Multi-task Model for Emotion and Offensive Aided Stance Detection of Climate Change Tweets

Apoorva Upadhyaya
upadhyaya@l3s.de
L3S Research Center
Hannover, Germany

Marco Fisichella*
mfisichella@l3s.de
L3S Research Center
Hannover, Germany

Wolfgang Nejdl
nejdl@l3s.de
L3S Research Center
Hannover, Germany

ABSTRACT

In this work, we address the United Nations Sustainable Development Goal 13: Climate Action by focusing on identifying public attitudes toward climate change on social media platforms such as Twitter. Climate change is threatening the health of the planet and humanity. Public engagement is critical to address climate change. However, climate change conversations on Twitter tend to polarize beliefs, leading to misinformation and fake news that influence public attitudes, often dividing them into climate change believers and deniers. Our paper proposes an approach to classify the attitude of climate change tweets (believe/deny/ambiguous) to identify denier statements on Twitter. Most existing approaches for detecting stances and classifying climate change tweets either overlook deniers' tweets or do not have a suitable architecture. The relevant literature suggests that emotions and higher levels of toxicity are prevalent in climate change Twitter conversations, leading to a delay in appropriate climate action. Therefore, our work focuses on learning stance detection (main task) while exploiting the auxiliary tasks of recognizing emotions and offensive utterances. We propose a multimodal multitasking framework MEMOCLiC that captures the input data using different embedding techniques and attention frameworks, and then incorporates the learned emotional and offensive expressions to obtain an overall representation of the features relevant to the stance of the input tweet. Extensive experiments conducted on a novel curated climate change dataset and two benchmark stance detection datasets (SemEval-2016 and ClimateStance-2022) demonstrate the effectiveness of our approach.

CCS CONCEPTS

• **Computing methodologies** → **Multi-task learning**; *Supervised learning by classification*; • **Human-centered computing** → *Social media.*

KEYWORDS

climate change, Twitter, stance detection, emotion recognition, offensive language

*Corresponding author

WWW '23, April 30–May 04, 2023, Austin, TX, USA

ACM ISBN 978-1-4503-9416-1/23/04…$15.00
https://doi.org/10.1145/3543507.3583860

ACM Reference Format:
Apoorva Upadhyaya, Marco Fisichella, and Wolfgang Nejdl. 2023. A Multi-task Model for Emotion and Offensive Aided Stance Detection of Climate Change Tweets. In *Proceedings of the ACM Web Conference 2023 (WWW '23), April 30–May 04, 2023, Austin, TX, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3543507.3583860

1 INTRODUCTION

Climate change is a major crisis for humanity, species, and the planet's ecosystem. According to the latest report of the Intergovernmental Panel on Climate Change (IPCC), inaction on climate change will threaten our well-being [33]. Although scientific evidence supports this issue, the public remains divided. For any climate action to be taken, the role of the public and their perception is crucial. Social media platforms like Twitter are popular and help raise public awareness about the climate crisis. However, these climate change conversations polarize beliefs and create opinion-based ideologies, resulting in bias, misinformation, and fake news that affect public attitudes toward climate change [16, 53]. As reported by USA Today[1], climate change hoaxes and falsehoods remain widespread on social media platforms like Twitter and Facebook, and credible warnings are lacking. Therefore, it is essential for governments and researchers to monitor climate change deniers' tweets to identify them and intervene. Identifying such content and understanding public attitudes towards climate change motivated us to use stance detection.

Stance Detection determines the author's perspective towards a target (for/against/neutral). To prevent missing any denier tweets that could be harmful if disseminated, we perform statement-level climate change stance detection. Climate-specific tweets have been used to identify attitudes, but either suffer from sarcasm hidden in the tweets or lack advanced architectures to focus primarily on stance detection [6, 17, 42, 44]. Several studies classified tweets into supporters or opponents of the target using the SemEval-2016 benchmark dataset, which contains 5 target topics, including climate change (364 tweets) [12, 47]. However, the limited number of climate tweets prevents these techniques from focusing on the specific characteristics of climate-specific tweets. Considering this, our focus is on developing a model that enhances stance detection by incorporating other modalities and auxiliary tasks.

Researchers found that the emotional content of tweets impacts climate change discussion [9, 19]. Further, social media is filled with offensive content that leads to harassment, cyberbullying, and flamestorms [7]. Conversations about climate change aren't any different [37], where negativity and toxicity prevail. According

[1]https://usatoday.com/story/tech/2022/01/21/climate-change-misinformation-facebook-youtube-twitter/6594691001/

to a recent study[2] by the Center for Countering Digital Hate, 10 publishers with 186 million subscribers are responsible for most climate-denying content on social media. As a consequence, we investigate the role of emotion recognition and offensive language identification as auxiliary tasks to the main task of detecting stances on climate change in tweets. We also focus on multimodal inputs, i.e., text and nonverbal cues (emojis in tweets), to build reliable classification models that help to identify the emotional state, the sentiment of the tweeter and the sarcasm hidden in it, which in turn helps to identify the correct tweet's stance [5, 10]. Here are examples of climate change tweets: ***(i)"Deny stance":*** *You can't even wipe your own ass, old man. Who writes this shit anyway?#ClimateHoax; What kind of liberal bull shit is this? wTF #ClimateHoax.* ***(ii)"Believe stance":*** *#ClimateEmergency the time to act is now #ClimateAction ..; let us not waste any more time to #ActOnClimate.* The deniers' tweets display anger, disgust, and negative emotions and contain toxic content that is either directly directed at groups or individuals in the form of insults or contains abusive and insulting language. In the second example, believers' tweets have a lighter tone, with motivational and expectant feelings. Positive emotions and non-toxic content are more common in tweets from believers than in tweets from deniers (as explained in Section 3). To identify the underlying stance of the tweet, we examine this association of different emotional and offensive expressions.

The main contributions of our proposed work are summarized below: ***(i).*** We create a new dataset of tweets with text, emojis, and annotations of stance, emotion, and offensive categories for the climate change domain, which will be useful for further research (The code and dataset are available here[3]). ***(ii).*** To our knowledge, this is the first cross-sectional study to use emotion recognition and offensive language identification as auxiliary tasks. ***(iii).*** We propose a multitasking system MEMOCLiC (Multi-task model for EMotion and Offensive aided stance detection of Climate Change Tweets) that focuses on learning stance detection (primary) while using emotion recognition (secondary) and offensive language identification (secondary) tasks. A variety of embedding techniques and modality attention frameworks are integrated into the proposed approach to capture the appropriate modality-specific features and the deep contextual interactions between them. Our integration module fuses learned emotion and offensive task features with stance features to obtain an overall representation of stance. ***(iv).*** We compare our proposed approach with state-of-the-art methods on our climate change dataset and on two benchmark datasets (SemEval-2016 and ClimateStance-2022). Experimental results show that the proposed framework improves the performance of the primary task, i.e., stance detection, by benefiting from the auxiliary tasks, i.e., emotion and offensive language identification, compared to its single- and multi-task variants and SOTA approaches.

2 RELATED WORK

Climate Change and Stance Detection Social media platforms have been used to hold discussions and disseminate information about climate change [1, 30]. Recently, [43] analyses the behavior of students on social media platforms related to climate change. The debates about climate change on Twitter, however, have become an extremely polarising issue, dividing opinion between climate change deniers and believers [16]. Since, climate change denial often leads to the spread of misinformation [53], the need to identify such tweets has become one of the most important tasks for society. Hence, our work focuses on classifying the stance of climate change tweets to identify public attitudes. Climate-specific studies either have focused on uncovering the effects of polarization in climate change tweets [50] or have identified the stances of polarised users or statements [6, 41]. Some of the recent works have focused on identifying the stances of climate-specific tweets. [42] proposed a multi-task framework with sentiment analysis as an auxiliary task for stance detection, but suffered from the drawback of identifying sarcasm in tweets, which motivates us to focus on using multi-modality in the form of emojis, which are able to detect the hidden sarcasm in tweets for other analysis tasks [5] as well as using the finer-grained emotions instead of three sentiment classes to evaluate the attitude of the tweet. The [44] has recently developed the ClimateStance dataset and presents reliable results with the basic architecture using BERT models. Other climate-specific works identifying deniers' tweets from climate change data also lack advanced architectures [6, 17]. Hence, these studies motivated us to develop an efficient model with better embedding techniques and architecture that can classify the attitude of a tweet on climate change. Although stance detection has been studied in several works on the popular SemEVAL-2016 [12, 46, 47] dataset, these previous studies did not focus on understanding the characteristics of climate change denier tweets because their number is relatively small (29 denier tweets) and the presence of the toxicity content that can be beneficial for identifying different stances in the dataset towards the target domains has also not been utilised. Therefore, our approach can perform stance detection while leveraging the tasks of identifying emotions and offensive expressions.

Emotion Recognition Previous literature has utilised the emotion and sentiment tasks for a multi-task architecture [36]. Some works on stance detection have stressed sentiment's importance [48], while others argue that sentiment undermines its performance [38]. However, several works have focused on the emotional aspects of climate change conversations and justified their role in climate change [19, 26]. Hence, these studies motivated us to investigate the role of emotion in classifying climate tweets.

Offensive Language Identification There is a plethora of research addressing various aspects of online toxicity such as classifying offensive posts [32], assessing their impact on online communities [24], predicting the triggers of toxicity [2], and detecting cyberbullying [15]. The climate change field[2] is no exception in this case either where higher levels of toxicity and negative emotions are found in the conversations [37]. There is also the possibility that online toxicity can lead to violent actions in the physical world as well, and therefore should be treated as a matter of serious social gravity [31]. This motivated us to investigate the impact of toxic content in climate change posts.

3 DATASET

Similar to previous works [21], we also use hashtags to collect a wider dataset by making use of hashtag quality to identify different

[2]https://mashable.com/article/toxic-ten-climate-denial-study
[3]https://github.com/apoorva-upadhyaya/Emotion_Offensive_Aided_Stance

Category	Anger	Anticipation	Disgust	Fear	Joy
Believe	9.99	**28.54**	4.80	16.28	**20.40**
Deny	**27.20**	9.86	**24.62**	20.88	1.53
Ambiguous	9.93	22.88	5.84	**24.08**	16.64
Category	**Sadness**	**Surprise**	**Trust**	**Positive**	**Negative**
Believe	8.21	10.32	**28.76**	**59.08**	15.79
Deny	**27.78**	**24.90**	14.27	5.27	**61.21**
Ambiguous	14.25	8.96	27.53	51.56	23.02

Table 1: % of emotions present in tweets

Category	Severe_Toxicity	Identity_Attack	Insult	Profanity
Believe	0.60	0.53	1.43	0.92
Deny	**6.70**	**1.82**	**17.62**	**10.06**
Ambiguous	1.33	0.59	2.62	1.70
Category	**Threat**	**Sexually_Explicit**	**Toxicity**	**Non_toxic**
Believe	2.19	0.32	0.42	**96.06**
Deny	**4.69**	**2.01**	**7.37**	78.35
Ambiguous	2.85	0.59	2.29	94.66

Table 2: % of offensive labels present in tweets

groups' stances. We select the denier and believer hashtags used by the previous works [41, 42]. We then collect real-time tweets from 28 July 2021 to 30 May 2022 using the Tweepy API[4] with the query hashtags. For both categories, we filter out the tweets that contain at least one emoji. Overall, we found 1,316 deniers and 11,423 believers tweets that contain emojis.

3.1 Data Annotation

Stance Detection (SD): Existing literature suggests that the presence of a hashtag indicating a stance does not guarantee that the tweet has the same stance [39]. Furthermore, removing query hashtags may cause a tweet to no longer express the same stance as it did with the query hashtag. Therefore, we remove the query hashtags from the tweets and perform manual annotation. Similar to [39, 44], we use favor, against, and ambiguous labels for the stance detection task. We classify each tweet into one of the three categories based on its stance on climate change: ***(i.) Favor (believe):*** The tweet suggests that climate change is real and happening (using terms that contain opinions and concern about climate change). ***(ii.) Against (deny):*** We can conclude from the tweet that the content is directed against climate change (expression of ignorance, opposition to climate action, government policy). ***(iii.)Ambiguous:*** The tweets do not express a clear stance towards climate change. Three trained annotators were accredited to annotate the tweets with appropriate stance labels (believe/deny/ambiguous). We first noted that the use of sarcasm in the tweets led to inconsistencies between annotators. However, the conflicting annotations were resolved using the emoji in the tweets, followed by appropriate discussions and mutual agreements between annotators. We calculated the inter-annotator agreement to check the quality of the annotations. We observed a Fleiss-Kappa [40] score of 0.81, indicating that the annotation and the presented dataset are of considerable quality. In total, we found **5,661**, **1,044**, and **2,176** tweets labelled with "believe", "deny", and "ambiguous" stances respectively, after manually annotating and deduplicating tweets based on their textual content.

Emotion Recognition (ER): To compute the emotions, we have the NRCLex[5] Python library. The library uses the NRC word-emotion association lexicon [23], which contains associations of words with eight emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, and trust) and two sentiments (negative and positive). We consider the emotion label corresponding to each tweet as a list of 10 elements. The preprocessed tweet is input to the NRCLex, which provides the top emotions of the tweet. The emotions present in the tweet are labelled 1, and the rest are labelled 0. This creates an emotion list with multiple labels for each tweet. Three trained annotators manually evaluated the labels for 1000 randomly selected tweets. We consider the final annotations generated after inter-annotator agreement as the ground truth (Fleiss-Kappa [40] score of 0.80), compared them to the annotations generated by NRCLex, and found an accuracy of 96.3%. To save time and cost, we consider the annotations provided by NRCLex for the emotion task. The percentage of each emotion is shown in Table 1. The presence of large negative emotions (as indicated in the table) in climate denier posts motivated us to examine toxic aspects between different categories of climate-related tweets.

Offensive Language Identification (OI): Previous works have used the Perspective API to detect various abuse and offensive categories within the textual content [11, 29, 32, 35]. We also used the Perspective API developed by Jigsaw and Google's Counter Abuse Technology team in Conversation-AI [14] to identify the different categories of abuse and toxicity in online conversations. The API returns a probability score between 0 and 1 for a total of 7 different offensive attributes/categories: *Severe_Toxicity, Identity_Attack, Insult, Profanity, Threat, Sexually_Explicit, and Toxicity* (more details on the categorization can be found here[6]). After a manual review and careful analysis of different thresholds (0.3, 0.5, 0.6, 0.7), we decided on a threshold of 0.5 to avoid missing any type of toxicity in the tweet. This means that if the attribute value is ≥ 0.5, we will consider the presence of the corresponding offending attribute/category for the particular tweet, otherwise, the tweet will be marked as *Non_toxic*. This generates an offensive label list of length 8 with multiple labels for each tweet. To evaluate the quality of the labels predicted by Perspective API, three trained annotators manually annotated 1000 randomly selected tweets from our dataset. Then we matched the annotations with the predicted labels for the same tweets. We found a Fleiss-Kappa [40] value of 0.78 between our manual annotations and the semi-supervised labels, indicating that the predicted labels are of considerable quality. The percentage of each offending attribute found in the tweets is shown in Table 2. The data **pre-processing** techniques and the **significance** of multimodality and auxiliary tasks are described in Appendix A.1.

4 METHODOLOGY

Problem Statement: Design a stance detection method that uses textual and emoji features and combines the emotional and offensive aspects to classify the attitude of a tweet on climate change into one of the polarized classes (believe/deny/ambiguous). *Please note that in the following sections, we abbreviate the task of stance detection as SD, emotion recognition with ER, and offensive*

[4]http://docs.tweepy.org/en/latest/streaming_how_to.html
[5]https://github.com/metalcorebear/NRCLex
[6]https://developers.perspectiveapi.com/

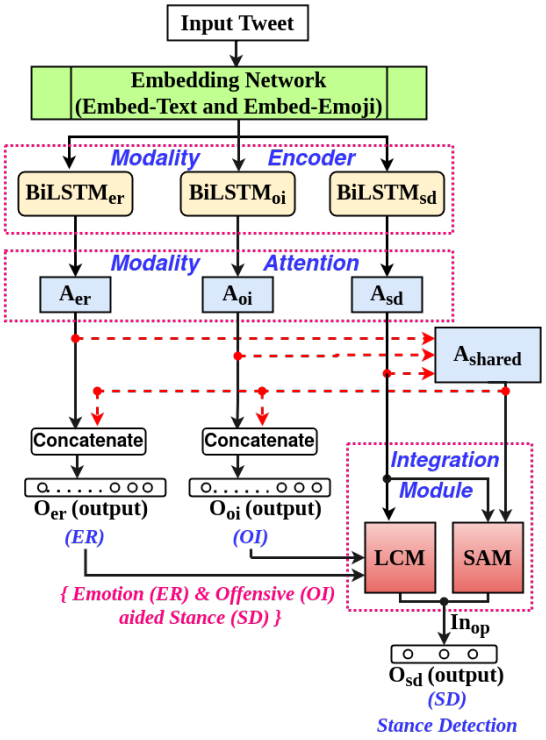


Figure 1: Architectural overview of our proposed MEMOCLiC framework

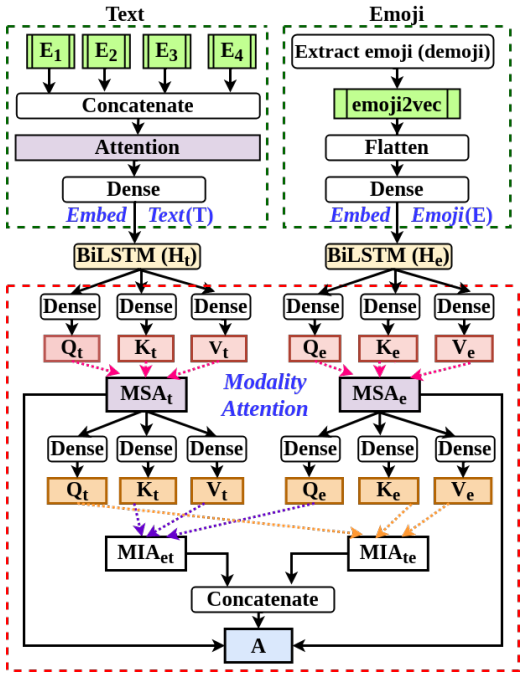


Figure 2: Overview of MEMOCLiC Components with text and emoji: (Top) Embedding Component (Embed-Text, Embed-Emoji) [common for all tasks]; (Bottom) Modality Attention [specific to each task]

language identification with OI. The proposed model MEMOCLiC consists of the following components: *Embedding Network, Modality Encoder, Modality Attention, Integration Module, and Classification Layer* (as shown in Figure 1). The input tweet (text and emoji) is embedded with the embedding network which is the same for all tasks. These features are then encoded with Bi-LSTM layers specific to each task ($BiLSTM_{sd}$,$BiLSTM_{er}$,$BiLSTM_{oi}$). The modality attention module provides task-specific features (A_{sd}, A_{er}, A_{oi}) that are averaged to obtain a shared attention vector (A_{shared}). The shared attention and the task-specific modality attention vectors are combined and fed into the softmax layer, resulting in the ER (O_{er}) and OI (O_{oi}) task outputs. These ER and OI outputs, along with the stance-specific attention and shared attention vectors, are passed through the integration module (In_{op}), followed by the softmax layer, which yields the final stance of the input tweet (O_{sd}). We now describe the components of the MEMOCLiC in detail.

4.1 Embedding Component

Embed-Text (T): The text representation of each input tweet is obtained through various embedding techniques. We use popular word embedding techniques such as GLOVE [20] and Pre-trained BERTweet (BERTweet) [27] that identify the semantics and syntax of each word in the text to create a vector representation of a tweet [13, 28]. However, these techniques mostly rely on the information from the neighboring words of a word. Therefore, to avoid the drawback of losing information by not capturing the semantics of the entire sentence, we have also applied sentence embeddings such as Sentence Bert (SBERT) [34] and Universal Sentence Encoder (USE) [4]. We refer to these embeddings as E_1: *GLOVE*, E_2: *BERTweet*, E_3: *SBERT, and* E_4: *USE.* Each tweet text 'T' containing n_t number of words, where the embedding of each word $w_1, ..w_{n_t}$ is obtained from a word embedding technique of dimension (d_{wt}), is flattened and results in ($T \in \mathbb{R}^{n_t(d_{wt})}$). The sentence embedding of a tweet text 'T' of dimension d_{st} results in $T \in \mathbb{R}^{d_{st}}$. We present two variants of the Embed-Text network based on the fusion of different embedding techniques to extract textual features:

(i)***Multi-Embedding Text-Concatenate (METC):*** In the METC variant, we concatenate the embeddings generated by the different word and sentence embedding techniques to efficiently capture the semantics and context of the input text so as to obtain a meaningful textual representation. Consequently, T = Concatenate (E_1, E_2, E_3, E_4) followed by a fully connected layer of dimension d_t yields the final text representation $T \in \mathbb{R}^{d_t}$, which is passed as an input to the next model component (*Modality Encoder*) and then used to train the multitasking framework for three tasks (refer Figure 1).

(ii)***Multi-Embedding Text-Attention (META):*** To focus on the most informative learned embeddings generated, we pass the output of the concatenated embeddings to the attention layer (using query, key, and value as detailed in Section 4.3) to obtain a final textual representation. Thus $T = Attention(Concatenate(E_1, E_2, E_3, E_4))$ followed by a fully connected layer of dimension d_t leads to the final textual representation as $T \in \mathbb{R}^{d_t}$. Figure 2 (top left) visually shows the META variant of the Embed Text component.

Embed-Emoji (E): The emoji features are extracted from tweets

using demoji[7], which is a Python library that extracts the image of the emoji from the text. Next, we use emoji2vec [8], which provides d_e = 300 dimensional vector representation for each of the emojis present in the tweet. If the input tweet contains n_e number of emojis, we obtain the final emoji representation after passing the vector from the emoji2vec to the flattened layer as E for a tweet, where $E \in \mathbb{R}^{n_e(d_e)}$ (refer Figure 2 (top right)).

4.2 Modality Encoders

The textual (T) and emoji (E) features obtained from the embedding component are then passed to the two discrete Bi-LSTM layers with dimension d_f to sequentially encode and learn complementary features based on semantic dependencies into hidden states for each of the modalities (as shown in Figure 2) and for each task separately (refer Figure 1). The hidden states of the Bi-LSTM layers provide a pair of the output of dimensions $H_t \in R^{2d_{tf}}$ and $H_e \in R^{2d_{ef}}$ for text and emoji respectively for each of the tasks. Hence, the output of the feature encoder is represented by $BiLSTM_{sd}$, $BiLSTM_{er}$, and $BiLSTM_{oi}$ for SD, ER and OI tasks respectively each containing a pair of vectors for text (H_t) and emoji (H_e) (refer Figure 1).

4.3 Modality Attention

Since the attention layer concentrates on the relevant part of the input and extracts the most important information from the input, we use the attention framework similarly to [45], in which the authors consider an attention function as a mapping to a set of queries, keys, and values. We pass the output of the Bi-LSTM layer of text (H_t) and emoji (H_e) through three fully connected layers of dimension d_a to obtain queries, keys, and values for the final feature representations. For our model, there are six triplets in total, forming three pairs of two triplets each for text (Q_t, K_t, V_t) and emoji (Q_e, K_e, V_e), which are used for SD (($Q_{tsd}, K_{tsd}, V_{tsd}$), ($Q_{esd}, K_{esd}, V_{esd}$)), ER (($Q_{ter}, K_{ter}, V_{ter}$), ($Q_{eer}, K_{eer}, V_{eer}$)), and OI (($Q_{toi}, K_{toi}, V_{toi}$), ($Q_{eoi}, K_{eoi}, V_{eoi}$)) tasks. Figure 2 (bottom) illustrates the following two attention frameworks used in our model:
Modality Specific Attention (MSA): Here, we relate different positions of an input sequence of the modality to identify the most important parts. We calculate the MSA scores using the equation 1 for text (MSA_t) and emoji (MSA_e) modalities for each of the tasks. The equation is shown in Figure 2 (bottom) as connections in pink dotted arrows. Here, three pairs of MSA scores are computed for SD (MSA_{sd}), ER (MSA_{er}), and OI (MSA_{oi}) tasks separately.

$$MSA_i = softmax(Q_i K_i^T)V_i \tag{1}$$

Modality Inter Attention (MIA): We find out the MIA scores to learn the interdependence between textual and emoji features. To obtain the query, key, and value for computing the MIA scores, we first pass the MSA scores of each modality to three fully connected layers of dimension d_a. MIA Scores are then determined using the following equations (2 and 3), intervening the query of one modality with the key and value of the other modality to reveal the significant contributions between these input modalities and learn optimal features for all tasks.

$$MIA_{et} = softmax(Q_e K_t^T)V_t, \tag{2}$$

$$MIA_{te} = softmax(Q_t K_e^T)V_e, \tag{3}$$

where $MIA_{te} \in \mathbb{R}^{d_a}$, and $MIA_{et} \in \mathbb{R}^{d_a}$. MIA equations are represented graphically with purple and brown dotted arrows in Figure 2 (bottom) part. The MSA and MIA scores are then concatenated, shown by $A = Concatenate(MSA_t, MSA_e, MIA_{et}, MIA_{te})$, where A_{sd}, A_{er}, and A_{oi} represents the attention output vector specific to each task of SD, ER and OI respectively (as shown in Figure 1). Furthermore, to take advantage of the shared features and use the features common to all tasks, we average the attention vector specific to each task, given by $A_{shared} = Average(A_{sd}, A_{er}, A_{oi})$ (refer Figure 1), which is then fed into the next component concatenated together with the task-specific attention output vector. The next component is the softmax layer for the ER (A_{er}, A_{shared}) and OI (A_{oi}, A_{shared}) tasks, while A_{sd} and A_{shared} are fed into the *Integration Module* for the SD task (see Figure 1).

4.4 Integration Module

This module is responsible for the fusion of emotions and offensive learned features to efficiently optimize the performance of stance detection. The module consists of the following 2 submodules:
Linear Convolution Module (LCM): is used to capture the overall attitude representation of an input tweet in terms of the associated emotion and toxicity levels present in the tweet. We take the outputs of the auxiliary tasks of ER ($O_{er} \in \mathbb{R}^{d_{oer}}$), OI ($O_{oi} \in \mathbb{R}^{d_{ooi}}$) and the final feature vector of stance-specific task ($A_{sd} \in \mathbb{R}^{d_a}$) as inputs to this module (where O_{er} and O_{oi} are obtained after passing the (A_{er}, A_{shared}) and (A_{oi}, A_{shared}) vectors through the softmax layer for ER and OI tasks respectively as shown in Figure 1). Since the existing literature suggests that the convolution operation efficiently models the effect of one function on the other and expresses how the shape of one is changed by the other [3, 49, 51], we also use the numpy operator convolve[8] to obtain the discrete linear convolution of A_{sd} with O_{er} and O_{oi} vectors. The equations 4 and 5 represent the convolution operations showing the effect of emotional and toxicity aspects onto the stance feature vector respectively:

$$(O_{er} * A_{sd})_n = \sum_{m=-\infty}^{\infty} O_{er}[m] A_{sd}[n-m], (LCM_1) \tag{4}$$

$$(O_{oi} * A_{sd})_n = \sum_{m=-\infty}^{\infty} O_{oi}[m] A_{sd}[n-m], (LCM_2) \tag{5}$$

$$LCM_{op} = Average(LCM_1, LCM_2) \tag{6}$$

where, n is the dimension of $O_{er}(R^{d_{oer}})$ in equation 4 and dimension of $O_{oi}(R^{d_{ooi}})$ in equation 5 and m is the dimension of $A_{sd}(R^{d_a})$. We finally average the output of the convolve function from the equations 4 and 5 as the final output vector of the LCM Module.
Stance Specific Shared Attention Module (SAM): Previous literature [52] has pointed out the disadvantage of multi-task learning, where there is a possibility that the shared space mixes some features irrelevant to the task, which makes task learning more difficult.

[7] https://pypi.org/project/demoji/

[8] https://numpy.org/doc/stable/reference/generated/numpy.convolve.html

To overcome this drawback, we apply a similar concept of Modality Inter Attention Framework (MIA), which uses query, key, and value to discard the useless shared features and focuses on the most informative shared features related to the stance detection task. The shared attention output vector (A_{shared}) is passed through the dense layers of dimension d_s to obtain the key (K_{shared}) and value (V_{shared}) of the shared features, while stance (A_{sd}) attention vector generates a query (Q_{sd}) for the stance-specific features. The query of the stance is intervened with the key and value of the shared features to focus on the relevant shared features with respect to the stance task (see equation 7).

$$SAM_{op} = softmax(Q_{sd}K^T_{shared})V_{shared} \quad (7)$$

Integration Cell: In various works [25], the fusion technique of absolute difference and element-wise product has been found to be effective, so equation 8 illustrates the final output of the integration cell as the fusion of LCM (LCM_{op}) and SAM (SAM_{op}) outputs.

$$In_{op} = [LCM_{op}; SAM_{op}; LCM_{op} - SAM_{op}; LCM_{op} \odot SAM_{op}] \quad (8)$$

4.5 Classification Layer

As depicted in Figure 1, the final predictions for ER and OI tasks are obtained by linearly concatenating the task-specific outputs (A_{er}, A_{oi}) with the shared output (A_{shared}) respectively. These outputs of the ER and the OI tasks are further used as input to the LCM of the integration module, as described in Section 4.4. The final output of the integration module (In_{op} as shown in the equation 8) is then fed into the softmax layer for the SD task. The integrated loss function (L) of our proposed system is realized as follows:

$$L = \alpha L_{sd} + \beta L_{er} + \gamma L_{oi} \quad (9)$$

We aggregate the weighted sum of the losses from the tasks to compute the overall loss (L_{sd} for SD, L_{er} for ER, and L_{oi} for OI). α, β, and γ represent the constants between 0 and 1 indicating the per-task loss-share to the overall loss.

5 EXPERIMENT

5.1 Dataset

We first conduct the experiments on our ***novel curated climate change dataset*** (details are covered in Section 3) followed by two benchmark stance detection datasets: ***(i.) ClimateStance-2022 [44]:*** is a recent publicly available dataset on climate change consisting of 3,777 tweets with "favor", "against" and "ambiguous" stances toward climate change prevention. The distribution of emotions and offensive content in the dataset can be found in Appendix A.2; ***(ii.) SemEval-2016 [22]:*** is a popular stance detection dataset used in SemEval-2016 shared task 6.A where tweets are in favor, against, or neutral corresponding to Atheism, Climate Change is a Real Concern, Feminist Movement, Hillary Clinton, and Abortion as targets. The distribution of emotion and offensive content in the dataset is given in Appendix A.2.

5.2 Experimental settings

We use the python-based library Keras (https://keras.io/) and Scikit-learn (https://scikit-learn.org/stable/) at various stages of our implementations. We consider accuracy, macro precision, macro recall, and macro F1 scores to evaluate the performance of our models.

Model	Single Task Stance Detection			
	Text		Text+Emoji	
	Macro F1	Accuracy	Macro F1	Accuracy
	Avg/St.dev	Avg/St.dev	Avg/St.dev	Avg/St.dev
GLOVE	74.09/0.46	75.64/0.86	76.08/1.11	77.47/0.37
GLOVE+MSA	76.27/1.12	78.31/1.16	79.88/0.67	80.66/0.21
BERTweet+MSA	78.23/0.58	80.01/0.83	81.17/1.32	83.38/1.09
SBERT+MSA	76.54/0.77	76.98/0.24	78.07/2.19	80.54/2.07
USE+MSA	77.31/1.08	78.29/0.81	80.93/1.47	82.82/1.34
METC+MSA	78.19/0.64	80.52/0.39	82.79/1.03	84.65/0.95
META +MSA	79.71/0.81	81.26/0.32	83.50/1.12	85.03/1.01
META +MA (MSA+MIA)	-	-	84.88/0.71	85.96/0.35
META +MA+ emo. and offens. as i/p features	81.79/0.38	84.03/1.31	85.78/0.87	87.66/1.22

Table 3: Results of the single task stance detection models in varying combinations

We perform stratified k-fold cross-validation on our dataset, oversample the minority classes (deny and ambiguous) in training data using sklearn resampling, and report averages and standard deviations (over 5 folds) for each metric. We run all the experiments on an NVIDIA GeForce GTX 1080Ti GPU. The best parameters of the reported results are as follows: Embeddings dimensions in *Embed-Text= GLOVE(E_1)*: 200, *BERTweet(E_2)*: 768, *SBERT(E_3)*: 768, *USE(E_4)*: 512, *Embed-Emoji (d_e)*: 300, *Bi-LSTM memory cells (d_f)*: 100, *fully connected layer dimension of modality attention (d_a) and SAM module (d_s) [with ReLu activation]*: 100, *Output dimension for ER (O_{er}) in LCM (d_{oer})*:10, *Output dimension for OI (O_{oi}) in LCM (d_{ooi})*:8, *output neurons/channels* : 3 [softmax activation] (SD), 10 [sigmoid activation] (ER) and 8 [sigmoid activation] (OI), *loss*: categorical cross-entropy (L_{sd}) for SD and binary cross-entropy loss function for ER (L_{er}) and OI (L_{oi}) tasks; *optimizer*: Adam(0.001). All the parameter values are selected using TPE in the Hyperopt[9] python library that minimises loss functions. Moreover, for the experiments, the loss weights for the SD *(α)*, ER *(β)*, and OI *(γ)* tasks are set as 1, 0.5 and 0.3 respectively. We fine-tune the loss weights for all tasks by using the Grid Search method from Scikit-learn.

5.3 Baseline models

We compare our proposed approach to the following baselines on our climate change dataset, which either detect climate change attitudes or classify tweets from diverse domains: ***RoBERTa-Base [44]***: performs stance detection on novel curated climate change tweets (ClimateStance dataset) with favor, against, and ambiguous labels. ***SP-MT [42]***: a novel multi-task framework that jointly performs stance detection and sentiment analysis on climate change Twitter dataset (believer and denier). ***MT-LRM-BERT [12]***: a multi-task framework that takes both sentiment and opinion-towards classification as auxiliary tasks for stance detection using SemEVAL-2016 and other benchmark datasets. ***S-MDMT [47]***: a multi-domain multi-task model to perform stance detection using SemEVAL-2016 dataset. ***ESD [46]***: performs stance detection by selecting an optimal ensemble of classifiers and feature set. ***HAN [48]***: a hierarchical attention neural model, focusing on the document, sentiment, dependency, and argument representations for stance

[9] http://hyperopt.github.io/hyperopt/

Model	Stance + Emotion (SD+ER)				Stance + Offensive (SD+OI)				Stance+Emotion+Offensive (SD+ER+OI)			
	Text		Text+Emoji		Text		Text+Emoji		Text		Text+Emoji	
	F1 score	Acc	F1 score	Acc	F1 score	Acc	F1 score	Acc	F1 score	Acc	F1 score	Acc
	Avg/St.dev	Avg/St.dev	Avg/St.dev	Avg/St.dev	Avg/St.dev	Avg/St.dev	Avg/St.dev	Avg/St.dev	Avg/St.dev	Avg/St.dev	Avg/St.dev	Avg/St.dev
METC+MA	82.66/1.30	83.98/1.42	84.54/1.08	86.33/1.27	81.66/1.09	82.99/2.12	84.06/1.05	85.55/0.64	84.54/1.23	85.19/1.01	86.51/1.05	89.45/1.07
META +MA	83.90/1.33	85.55/0.32	87.35/1.01	89.18/0.96	83.36/0.21	86.53/0.50	85.66/1.58	87.51/1.09	86.38/2.11	89.17/2.13	89.20/0.74	91.17/0.39
META+MA+LCM	87.01/0.58	88.10/1.04	90.31/1.41	92.51/2.09	87.12/0.63	89.72/0.69	88.89/0.22	90.63/1.09	88.34/1.41	92.10/1.72	92.04/0.72	94.73/0.38
META+MA+SAM	85.19/0.65	87.18/0.55	89.41/0.34	90.77/0.90	85.41/0.62	86.59/1.04	87.52/1.66	89.11/2.12	89.04/1.05	90.61/0.41	90.23/2.25	92.89/1.93
META+MA Integ.(LCM,SAM)	89.62/0.16	90.50/0.23	92.05/0.69	93.88/0.62	87.31/0.31	88.02/0.75	90.51/1.61	91.69/1.28	90.67/1.01	91.99/1.10	**93.76/0.62 (MEMOCLiC)**	**95.15/0.88 (MEMOCLiC)**

Table 4: Results of Stance Detection in Multi-task architectures on Climate Change Dataset (Macro F1 score & Accuracy). MEMOCLiC outperforms other variants while meeting statistical significance under t-tests (p <0.05).

Model	Precision	Recall	F1 score	Acc.
-	Avg/St.dev	Avg/St.dev	Avg/St.dev	Avg/St.dev
MEMOCLiC[Proposed]	**92.06/0.81**	**95.44/0.29**	**93.76/0.62**	**95.15/0.88**
RoBERTa-Base[44]	83.38/1.55	85.24/1.28	84.69/1.89	86.42/2.01
SP-MT[42]	87.95/1.11	90.01/1.80	89.29/1.31	91.47/0.91
MT-LRM-BERT [12]	87.12/1.61	88.70/0.99	88.59/1.29	90.01/1.67
S-MDMT [47]	86.12/1.02	88.67/0.39	86.91/0.44	88.33/0.48
ESD [46]	81.55/1.72	84.39/2.05	83.28/2.31	86.92/2.01
HAN[48]	84.61/1.22	84.23/1.78	84.54/1.65	86.59/1.82
MNB[17]	78.11/0.66	79.51/0.73	78.43/1.33	80.16/1.32
DNN[6]	77.64/1.58	76.38/1.08	77.15/1.18	79.92/1.34

Table 5: Results for Stance Detection on Climate Change Dataset with Baselines. MEMOCLiC outperforms all baselines while meeting statistical significance under t-tests (p <0.05).

detection. ***MNB [17]***: Multinomial naive bayes classifies tweets into positive, negative, or neutral beliefs towards climate change. ***DNN [6]***: a neural network to identify Twitter users as climate change deniers/believers.

6 RESULTS

To demonstrate the importance of the varying modalities and components of the proposed approach, we first compare the different single-task variants of MEMOCLiC. The MEMOCLiC framework is then analyzed with regard to a variety of multi-task variants, followed by a comparison with state-of-the-art algorithms on a variety of datasets. *It is to be noted that the current work aims to improve the performance of stance detection (SD) with the help of the other two secondary tasks (ER and OI). Therefore, we report the results and analysis with SD as the main task in all task combinations.*

Comparison amongst single-task stance detection frameworks on climate change dataset: It can be seen from Table 3, the addition of nonverbal cues in the form of emojis consistently improves the uni-modal textual baseline. This improvement means that the proposed architecture makes very effective use of the interaction between input modalities and highlights the importance of incorporating multi-modal features for the analysis tasks. The table also shows that the BERTweet embedding combined with modality specific attention (BERTweet+MSA) performs better compared to the other embeddings (GLOVE +MSA, SBERT+MSA, and USE +MSA) because the BERTweet embeddings are trained on tweets only and easily capture the textual features in the tweets, regardless of the drawback of the short length of the tweet. However, concatenating the different embeddings followed by the attention layer further improves the performance of the task with the F1 score of 83.50 (META +MSA). This 2.87% improvement in the average F1 score (compared to the BERTweet+MSA F1 score of 81.17) indicates that the model is able to learn efficiently, suggesting that the different embedding spaces provide additional information and can understand context, intent, and other nuances in the tweet text. In addition, the best performing embedding component (META) along with the attention frameworks of MSA and MIA improved the task by achieving the F1 score of 84.88. This shows that MSA and MIA capture the most important modality-specific and interactive features. Results also improved when we used emotion and toxicity as input features, supporting the argument that emotional and offensive features effectively enhance the learning of the attitude of a tweet. The corresponding precision and recall of the models are given in Table 12 in the Appendix B.

Comparison amongst different multi-task stance detection frameworks on climate change dataset: Table 4 shows the performance of the proposed approach for the different combinations of the stance detection task with other auxiliary tasks (Stance + Emotion [SD+ER], Stance + Offensive [SD+OI], and Stance + Emotion + Offensive [SD+ER+OI]). From Table 4, it can be observed that in the MEMOCLiC framework, the LCM & SAM submodules of the integration component improved the F1 score of the model (93.76) by 5.11% improvement over the META embedding with Modality Attention (MA) component (89.20), as the linear convolution function effectively captures the emotions and toxic aspects of the tweets in relation to the stance of the tweet, while the addition of SAM further improves the task by focusing on the useful shared features with respect to the stance task. Table 4 also suggests the better performance of the SD+ER combination than SD+OI with 92.05 and 90.51 mean F1 scores, respectively. This is because a clearer split in terms of emotion is visible between the three attitude categories of the tweet (Table 1) than the offending labels with a high proportion of non-toxic content within believe and ambiguous labels (Table 2). However, combining the two auxiliary tasks with the main stance detection task improves the overall performance of the framework with 93.76 F1 score. The improvement in F1 score from 85.78 of single-task stance detection using emotion and offensive as input features (see Table 3) to 93.76 using the MEMOCLiC framework validates the importance of both the auxiliary tasks and our proposed multitasking approach and its components for classifying the stance of a tweet. The precision and recall are in Appendix B.

Comparison with Baselines: *(i.) Climate Change Dataset:* The MEMOCLiC model outperforms the SOTA approaches (see Table 5). Emojis remove the drawback of the SP-MT model by processing sarcasm contained in tweets. The SP-MT and MT-LRM-BERT models perform better than the other baselines since they include sentiment and opinion formation tasks. MEMOCLiC outperforms these methods and demonstrates that adding different emotions

Model	Precision	Recall	F1 score	Acc.
MEMOCLiC[Proposed]	**0.554**	**0.5206**	**0.537**	81.49%
RoBERTa-Base[44]	0.528	0.502	0.510	81.22%
BERT-Base	0.507	0.446	0.464	77.51%
BERT-Large	0.530	0.470	0.489	77.78%
RoBERTa-Large	0.473	0.507	0.489	**82.54%**
DistilBERT	0.497	0.430	0.448	79.37%

Table 6: Results for Stance Detection on benchmark ClimateStance-2022 dataset

Model	Atheism F_{avg}	Climate F_{avg}	Feminism F_{avg}	Hillary F_{avg}	Abortion F_{avg}	Mac F_{avg}
MEMOCLiC[Proposed]	74.39	**64.51**	63.62	**75.84**	**71.36**	**69.94**
MT-LRM-BERT[12]	**76.14**	53.05	63.12	74.67	70.32	67.46
SP-MT[42]	69.5	63.5	63.2	67.5	70.5	66.84
S-MDMT[47]	69.50	52.49	**63.78**	67.20	67.19	64.03
ESD[46])	66.64	43.82	62.85	67.79	64.94	61.20
HAN[48]	70.53	49.56	57.50	61.23	66.16	61.00
AT-JSS-LEX[18]	69.22	59.18	61.49	68.33	68.41	65.33
SVM-ngram[39]	65.19	42.35	57.46	58.63	66.42	58.01

Table 7: Results for Stance Detection on SemEval-2016 Dataset with Baselines

and toxicity levels extracted from tweets with different embedding combinations and attention frameworks in a multitask setting enhances SD performance. With MEMOCLiC, task-specific and shared features were used to improve task performance, suggesting that the shared private approach with emotions and toxicity awareness is more effective than ESD and HAN. Using ER and OI as auxiliary tasks to support SD, the proposed approach MEMOCLiC outperforms the S-MDMT model implemented with a multitask approach with target classification as a separate task. A single-task SD model with an F1 score of 84.88 (Table 3) outperforms both DNN and MNB approaches with an average F1 score increase of 9.12%. Thus, by utilizing better embedding techniques and attention frameworks with different modalities, a better architecture can improve classification. ***(ii.) ClimateStance-2022 Dataset:*** We use the baseline methods from the [44] work that created the ClimateStance dataset. From Table 6, MEMOCLiC performs significantly better than the baselines with an average increase of macro F1 score of 12.10% in an imbalanced dataset, suggesting that emotion and toxicity aspects in the tweet with different combinations of embedding (BERTweet is also the relevant difference compared to the BERT models in the baseline methods) are effective in detecting the attitude of a climate change tweet. ***(iii.) SemEval-2016 Dataset:*** The metrics (F_{avg}, $MacF_{avg}$) are calculated according to the approach specified in the work [18]. Based on Table 7, it is observed that MEMOCLiC outperforms all other baselines on the benchmark dataset, especially for the *climate, Hillary, and abortion* targets, with $MacF_{avg}$ of 69.94 as it provides better discrimination and clearer separation in terms of emotion and offensive characteristics between the favor and against classes in the dataset (Appendix A.2). The *atheism* and *feminism* targets also have comparable F_{avg} scores. As a result, MEMOCLiC can also be applied to different domains and topics, which proves the generalisability of our approach.

6.1 Error Analysis

We discuss possible reasons for the errors in SD task: ***(i.) Skewness of Dataset:*** The skewed class distribution of the climate change dataset affects the predictions of the MEMOCLiC model (*deny:* 11.76%, *ambiguous:* 24.50%, and *believe:* 63.74%). While we applied oversampling to partially address this issue, further categorization of believers can aid in balancing data for the different groups. ***(ii.) Close proximity between believe and ambiguous:*** The tweets like *"The only way forward before we cross the boundary of no turning back. In fact, we might already ...."* labelled as "ambiguous," depending on what the tweet text conveys. However, since the emotion of expectation along with the non-toxic content predominates in the believe tweet category, our model predicts the incorrect stance as "believe". ***(iii.) Composite Tweets:*** Example tweets in the dataset contain multiple sentences that cover contrasting emotions, making predicting the correct label stance difficult. For example, *Glad you're are back... #Cop26 is the biggest #con known to man along with...; predicted class: believe (incorrect).* Positive and negative emotions are conveyed in the first sentence. Due to the composite nature of the tweet with conflicting emotions, the model makes incorrect predictions since it focuses on the emotions in the first sentence. These scenarios limit the performance of MEMOCLiC. However, in our future work, we will focus on the causal extraction behind the emotion and toxicity levels in the tweets, extracting unique features for stance categories and other modalities, such as images and videos that can provide a better insight into stance of the tweet.

7 CONCLUSION

In this paper, we address "climate action", one of the United Nations Sustainable Development Goals. Because our work is dedicated to classifying the stance of a tweet (believe/deny/ambiguous), our proposed approach will be useful for the government and researchers to combat climate misinformation by identifying posts from the climate change deniers and reducing their spread. We curate a novel climate change dataset consisting of different modalities in the form of text and emojis with annotations for stance, emotion, and offensive categories, which is beneficial for the research community. We propose a multitasking model that uses the learned features of emotion (auxiliary) and offensive tasks (auxiliary) to optimize stance detection (primary). The results of the experiments conducted on a novel curated and two benchmark datasets show that multi-modality and multi-tasking increase the performance of the stance task compared to its single-task variants and baseline methods by leveraging the auxiliary tasks. It is also observed that the model is much more broadly applicable beyond the climate change domain based on its performance on the SemEval dataset, suggesting the generalisability of the proposed approach. In the future, we will attempt to focus on the annotation of composite tweets using other NLP tasks such as aspect-based sentiment, and other modalities such as images and videos to predict the more accurate classification of polarized attitudes toward climate change.

ACKNOWLEDGMENTS

This work was partly funded by the SoMeCliCS project under the Volkswagen Stiftung and Niedersächsisches Ministerium für Wissenschaft und Kultur and by the European Commission for the eXplainable Artificial Intelligence in healthcare Management (xAIM) project, agreement no. INEA/CEF/ICT/A2020/2276680.

REFERENCES

[1] Joachim Allgaier. 2019. Science and environmental communication on YouTube: strategically distorted communications in online videos on climate change and climate engineering. *Frontiers in Communication* 4 (2019), 36.

[2] Hind Almerekhi, Haewoon Kwak, Joni Salminen, and Bernard J Jansen. 2020. Are these comments triggering? predicting triggers of toxicity in online discussions. In *Proceedings of The Web Conference 2020*. 3033–3040.

[3] Mawardi Bahri, Ryuichi Ashino, and Rémi Vaillancourt. 2013. Convolution theorems for quaternion Fourier transform: properties and applications. In *Abstract and Applied Analysis*, Vol. 2013. Hindawi.

[4] Daniel Cer, Yinfei Yang, Sheng-yi Kong, Nan Hua, Nicole Limtiaco, Rhomni St John, Noah Constant, Mario Guajardo-Cespedes, Steve Yuan, Chris Tar, et al. 2018. Universal sentence encoder for English. In *Proceedings of the 2018 conference on empirical methods in natural language processing: system demonstrations*. 169–174.

[5] Dushyant Singh Chauhan, Gopendra Vikram Singh, Aseem Arora, Asif Ekbal, and Pushpak Bhattacharyya. 2022. An emoji-aware multitask framework for multimodal sarcasm detection. *Knowledge-Based Systems* 257 (2022), 109924.

[6] Xingyu Chen, Lei Zou, and Bo Zhao. 2019. Detecting climate change deniers on twitter using a deep neural network. In *Proceedings of the 2019 11th International Conference on Machine Learning and Computing*. 204–210.

[7] Justin Cheng, Michael Bernstein, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. 2017. Anyone can become a troll: Causes of trolling behavior in online discussions. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*. 1217–1230.

[8] Ben Eisner, Tim Rocktäschel, Isabelle Augenstein, Matko Bošnjak, and Sebastian Riedel. 2016. emoji2vec: Learning Emoji Representations from their Description. In *Proceedings of the Fourth International Workshop on Natural Language Processing for Social Media*. Association for Computational Linguistics. https://doi.org/10.18653/v1/W16-6208

[9] May El Barachi, Manar AlKhatib, Sujith Mathew, and Farhad Oroumchian. 2021. A novel sentiment analysis framework for monitoring the evolving public opinion in real-time: Case study on climate change. *Journal of Cleaner Production* (2021), 127820.

[10] Bjarke Felbo, Alan Mislove, Anders Søgaard, Iyad Rahwan, and Sune Lehmann. 2017. Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, Copenhagen, Denmark, September 9-11, 2017*, Martha Palmer, Rebecca Hwa, and Sebastian Riedel (Eds.). Association for Computational Linguistics, 1615–1625. https://doi.org/10.18653/v1/d17-1169

[11] Paula Fortuna, Juan Soler, and Leo Wanner. 2020. Toxic, hateful, offensive or abusive? what are we really classifying? an empirical analysis of hate speech datasets. In *Proceedings of the 12th language resources and evaluation conference*. 6786–6794.

[12] Yujie Fu, Xiaoli Li, Yang Li, Suge Wang, Deyu Li, Jian Liao, and Jianxing Zheng. 2022. Incorporate opinion-towards for stance detection. *Knowledge-Based Systems* 246 (2022), 108657.

[13] Soumitra Ghosh, Asif Ekbal, and Pushpak Bhattacharyya. 2022. Deep cascaded multitask framework for detection of temporal orientation, sentiment and emotion from suicide notes. *Scientific reports* 12, 1 (2022), 1–16.

[14] Hossein Hosseini, Sreeram Kannan, Baosen Zhang, and Radha Poovendran. 2017. Deceiving google's perspective api built for detecting toxic comments. *arXiv preprint arXiv:1702.08138* (2017).

[15] Homa Hosseinmardi, Sabrina Arredondo Mattson, Rahat Ibn Rafiq, Richard Han, Qin Lv, and Shivakant Mishra. 2015. Analyzing labeled cyberbullying incidents on the instagram social network. In *International conference on social informatics*. Springer, 49–66.

[16] S Mo Jang and P Sol Hart. 2015. Polarized frames on "climate change" and "global warming" across countries and states: Evidence from Twitter big data. *Global Environmental Change* 32 (2015), 11–17.

[17] Chuma Kabaghe and Jason Qin. 2020. Classifying tweets based on climate change stance. *Training* 66, 60.9 (2020), 61.

[18] Yingjie Li and Cornelia Caragea. 2019. Multi-task stance detection with sentiment and stance lexicons. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. 6299–6305.

[19] Maria L Loureiro and Maria Alló. 2020. Sensing climate change and energy issues: Sentiment and emotion analysis with social media in the UK and Spain. *Energy Policy* 143 (2020), 111490.

[20] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* (2013).

[21] Amita Misra, Brian Ecker, Theodore Handleman, Nicolas Hahn, and Marilyn Walker. 2016. NLDS-UCSC at SemEval-2016 Task 6: A Semi-Supervised Approach to Detecting Stance in Tweets. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*. Association for Computational Linguistics.

[22] Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016. Semeval-2016 task 6: Detecting stance in tweets. In *Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016)*. 31–41.

[23] Saif M Mohammad and Peter D Turney. 2013. Crowdsourcing a word–emotion association lexicon. *Computational intelligence* 29 (2013).

[24] Shruthi Mohan, Apala Guha, Michael Harris, Fred Popowich, Ashley Schuster, and Chris Priebe. 2017. The impact of toxic language on the health of reddit communities. In *Canadian Conference on Artificial Intelligence*. Springer, 51–56.

[25] Lili Mou, Rui Men, Ge Li, Yan Xu, Lu Zhang, Rui Yan, and Zhi Jin. 2015. Natural language inference by tree-based convolution and heuristic matching. *arXiv preprint arXiv:1512.08422* (2015).

[26] Robin L Nabi, Abel Gustafson, and Risa Jensen. 2018. Framing climate change: Exploring the role of emotion in generating advocacy behavior. *Science Communication* 40, 4 (2018), 442–468.

[27] Dat Quoc Nguyen, Thanh Vu, and Anh Tuan Nguyen. 2020. BERTweet: A pre-trained language model for English Tweets. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. Association for Computational Linguistics, Online. https://doi.org/10.18653/v1/2020.emnlp-demos.2

[28] Thi Huyen Nguyen and Koustav Rudra. 2022. Towards an Interpretable Approach to Classify and Summarize Crisis Events from Microblogs. In *Proceedings of the ACM Web Conference 2022*. 3641–3650.

[29] Adewale Obadimu, Esther Mead, Muhammad Nihal Hussain, and Nitin Agarwal. 2019. Identifying toxicity within youtube video comment. In *International conference on social computing, Behavioral-cultural modeling and prediction and behavior representation in modeling and simulation*. Springer, 214–223.

[30] Neetu Pathak, Michael J. Henry, and Svitlana Volkova. 2017. Understanding Social Media's Take on Climate Change through Large-Scale Analysis of Targeted Opinions and Emotions. In *2017 AAAI Spring Symposia, Stanford University, Palo Alto, California, USA, March 27-29, 2017*. AAAI Press. http://aaai.org/ocs/index.php/SSS/SSS17/paper/view/15341

[31] Desmond Upton Patton, Robert D Eschmann, Caitlin Elsaesser, and Eddie Bocanegra. 2016. Sticks, stones and Facebook accounts: What violence outreach workers know about social media and urban-based gang violence in Chicago. *Computers in human behavior* 65 (2016), 591–600.

[32] John Pavlopoulos, Nithum Thain, Lucas Dixon, and Ion Androutsopoulos. 2019. Convai at semeval-2019 task 6: Offensive language identification and categorization with perspective and bert. In *Proceedings of the 13th international Workshop on Semantic Evaluation*. 571–576.

[33] Hans O Pörtner, Debra C Roberts, Helen Adams, Carolina Adler, Paulina Aldunce, Elham Ali, Rawshan Ara Begum, Richard Betts, Rachel Bezner Kerr, Robbert Biesbroek, et al. 2022. Climate change 2022: impacts, adaptation and vulnerability. (2022).

[34] Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics. https://arxiv.org/abs/1908.10084

[35] Sayar Ghosh Roy, Ujwal Narayan, Tathagata Raha, Zubair Abid, and Vasudeva Varma. 2020. Leveraging Multilingual Transformers for Hate Speech Detection. In *Working Notes of FIRE 2020 - Forum for Information Retrieval Evaluation, Hyderabad, India, December 16-20, 2020 (CEUR Workshop Proceedings, Vol. 2826)*. CEUR-WS.org, 128–138.

[36] Tulika Saha, Apoorva Upadhyaya, Sriparna Saha, and Pushpak Bhattacharyya. 2021. A Multitask Multimodal Ensemble Model for Sentiment-and Emotion-Aided Tweet Act Classification. *IEEE Transactions on Computational Social Systems* (2021).

[37] Mary Sanford, James Painter, Taha Yasseri, and Jamie Lorimer. 2021. Controversy around climate change reports: a case study of Twitter responses to the 2019 IPCC report on land. *Climatic change* 167, 3 (2021), 1–25.

[38] Indira Sen, Fabian Flöck, and Claudia Wagner. 2020. On the reliability and validity of detecting approval of political actors in tweets. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 1413–1426.

[39] Parinaz Sobhani, Saif Mohammad, and Svetlana Kiritchenko. 2016. Detecting stance in tweets and analyzing its interaction with sentiment. In *Proceedings of the fifth joint conference on lexical and computational semantics*. 159–169.

[40] Robert L Spitzer, Jacob Cohen, Joseph L Fleiss, and Jean Endicott. 1967. Quantification of agreement in psychiatric diagnosis: A new approach. *Archives of General Psychiatry* 17, 1 (1967), 83–87.

[41] Aman Tyagi, Matthew Babcock, Kathleen M Carley, and Douglas C Sicker. 2020. Polarizing tweets on climate change. In *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation*. Springer, 107–117.

[42] Apoorva Upadhyaya, Marco Fisichella, and Wolfgang Nejdl. 2022. A Multi-task Model for Sentiment Aided Stance Detection of Climate Change Tweets. *arXiv preprint arXiv:2211.03533* (2022).

[43] Apoorva Upadhyaya, Catharina Pfeiffer, Oleh Astappiev, Ivana Marenzi, Stefanie Lenzer, Andreas Nehring, and Marco Fisichella. 2022. How learnweb can support

science education research on climate change in social media. In *12th International Conference on Methodologies and Intelligent Systems for Technology Enhanced Learning*. Springer.
[44] Roopal Vaid, Kartikey Pant, and Manish Shrivastava. 2022. Towards Fine-grained Classification of Climate Change related Social Media Text. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*. 434–443.
[45] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in neural information processing systems*. 5998–6008.
[46] Sergey Vychegzhanin and Evgeny Kotelnikov. 2021. A New Method for Stance Detection Based on Feature Selection Techniques and Ensembles of Classifiers. *IEEE Access* 9 (2021), 134899–134915.
[47] Limin Wang and Dexin Wang. 2021. Solving Stance Detection on Tweets as Multi-Domain and Multi-Task Text Classification. *IEEE Access* 9 (2021), 157780–157789.
[48] Zhongqing Wang, Qingying Sun, Shoushan Li, Qiaoming Zhu, and Guodong Zhou. 2020. Neural Stance Detection With Hierarchical Linguistic Representations. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 28 (2020), 635–645.
[49] Wikipedia. 2013. "Convolution". https://en.wikipedia.org/wiki/Convolution
[50] Hywel TP Williams, James R McMurray, Tim Kurz, and F Hugo Lambert. 2015. Network analysis reveals open forums and echo chambers in social media discussions of climate change. *Global environmental change* 32 (2015), 126–138.
[51] Jiaze Wu, Hao Liang, Changsong Ding, Xindi Huang, Jianhua Huang, and Qinghua Peng. 2021. Improving the accuracy in classification of blood pressure from photoplethysmography using continuous wavelet transform and deep learning. *International journal of hypertension* 2021 (2021).
[52] Lianwei Wu, Yuan Rao, Haolin Jin, Ambreen Nazir, and Ling Sun. 2019. Different absorption from the same sharing: Sifted multi-task learning for fake news detection. In *Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. 4636–4645.
[53] Yanmengqian Zhou and Lijiang Shen. 2021. Confirmation Bias and the Persistence of Misinformation on Climate Change. *Communication Research* (2021), 00936502211028049.

A DATASET

We select the hashtags of denier (*ClimateHoax, YellowVests and Qanon*) and believer (*ClimateChangeIsReal, ClimateActionNow, FactsMatter, ScienceMatters, ScienceIsReal*) as query hashtags used in previous works [41, 42] to collect real-time tweets. *Please note that we share our dataset only with the tweet ids and annotations. Though we conduct our research using Twitter data, but we strictly adhere to privacy protections, so we do not provide personally identifiable information.*

A.1 Data pre-processing

As described in Section 3.1, we first remove the query hashtags from the tweets for the manual annotation process, as it is possible that the stance of the tweet gets changed after removing the query hashtags. We further preprocess the tweet text by removing mentions, URLs, punctuation, spaces, stopwords, and unwanted characters such as RT and CC. All words are converted to lowercase. Then, we use an NLTK-based tokenizer to tokenize tweets. We also reduce the inflected words by applying NLTK Wordnet Lemmatizer.

Significance of Multi-modality In Figure 3, we present examples from the dataset to highlight the importance of nonverbal cues such as emojis in a tweet along with the text. Examples of tweets in the deny stance category (ex. 1, 2, 3): The emotions in the tweets are mostly negative and are accompanied by anger or disgust. However, the emojis show either a laughing or a thoughtful face, suggesting that the emojis contain complementary information that helps to reveal the sarcasm in the deniers' tweets, thus solving the problem of identifying the attitude of a tweet based on the sarcasm hidden in it faced by previous works. Moreover, the believers' tweets are mostly tagged with the same emotions and more positive emojis, suggesting that text and emojis contain similar information for the believers' category, which further facilitates the stance detection task by allowing distinction between deny and believe stance categories of tweets.

Category	Anger	Anticipation	Disgust	Fear	Joy
Favor	8.90	15.78	3.79	30.89	8.69
Against	15.89	15.14	12.01	48.06	8.52
Ambiguous	11.03	14.38	6.66	31.14	9.36
Category	**Sadness**	**Surprise**	**Trust**	**Positive**	**Negative**
Favor	6.96	5.85	21.47	53.05	29.47
Against	14.34	12.79	18.21	31.47	47.28
Ambiguous	7.69	6.68	18.39	42.14	31.43

Table 8: % of Emotions present in different stances of ClimateStance-2022 Dataset

Category	Severe_Toxicity	Identity_Attack	Insult	Profanity
Favor	1.87	1.27	5.05	2.86
Against	3.76	2.51	17.55	6.89
Ambiguous	3.13	2.08	8.09	5.22
Category	**Threat**	**Sexually_Explicit**	**Toxicity**	**Non_toxic**
Favor	2.65	0.9	2.22	92.18
Against	4.70	2.19	8.46	79.62
Ambiguous	4.69	1.82	4.96	86.68

Table 9: % of Offensive Expressions present in different stances of ClimateStance-2022 Dataset

Category	Anger	Anticipation	Disgust	Fear	Joy
Favor	13.26	17.56	6.45	27.24	11.82
Against	8.33	25	12.5	33.33	12.5
Neutral	12.65	20.48	7.23	20.48	20.49
Category	**Sadness**	**Surprise**	**Trust**	**Positive**	**Negative**
Favor	13.26	6.09	29.03	54.48	36.55
Against	33.33	8.33	25	16.66	79.16
None	11.44	7.22	29.51	59.63	32.53

Table 10: % of Emotions present in different stances of "Climate Change is a Real Concern" target of SemEval-2016 Dataset

Category	Severe_Toxicity	Identity_Attack	Insult	Profanity
Favor	6.28	7.48	8.08	7.78
Against	11.54	19.23	15.38	3.84
None	9.85	13.79	11.33	8.37
Category	**Threat**	**Sexually_Explicit**	**Toxicity**	**Non_toxic**
Favor	9.58	7.78	6.69	81.73
Against	15.38	7.69	11.68	73.07
None	14.77	14.28	9.90	74.38

Table 11: % of Offensive Expressions present in different stances of "Climate Change is a Real Concern" target of SemEval-2016 Dataset

Significance of Emotions & Offensive Features From Figure 3, examples 1 and 3 of the denial category include words such as "solar panels, charge your vehicle, carbon footprint," which are mostly present in tweets with a believer attitude, which could confuse

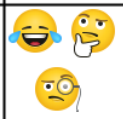
S.No.	Stance	Tweet	Emoji	Emotions	Offensive labels
1.	Deny	charge you vehicle at night....from your solar panels...did he really say this..and they think We are dumb.		surprise, negative	insult
2.	Deny	Fools and morons backing the #ClimateHoax religion of the socialist party.		disgust, negative	severe_toxicity, identity_attack, insult, toxicity
3.	Deny	you own three houses asshole... clean up your own stupid ""carbon footprint "" first #ClimateChangeHoax"		anger,disgust, negative	severe_toxicity, insult, profanity ,toxicity
4.	Believe	wonderful they provide #veganfood @fflglobal #plantbasedfood this is called #ClimateAction		trust, joy positive	non-toxic
5.	Believe	Why not own and nurture a tree today...Imagine having a life share life even after you are gone #PetaTree#ClimateEmergency		anticipation, positive	non-toxic

Figure 3: Significance of incorporating emoji, emotion and offensive expressions for stance detection

Model	Single Task Stance Detection			
	Text		Text+Emoji	
	Precision	Recall	Precision	Recall
	Avg/St.dev	Avg/St.dev	Avg/St.dev	Avg/St.dev
GLOVE	75.52/0.11	74.26/0.18	76.50/0.14	75.92/0.91
GLOVE+MSA	76.59/1.41	75.92/2.05	80.16/0.55	79.45/0.71
BERTweet+MSA	77.41/0.39	80.16/0.65	81.24/1.53	83.49/1.50
SBERT+MSA	77.09/1.33	76.8/0.12	79.07/2.06	81.47/1.45
USE+MSA	75.61/1.39	76.90/0.48	80.36/0.12	82.02/2.15
METC+MSA	78.25/0.41	79.52/0.39	81.02/1.04	84.24/0.76
META+MSA	79.08/1.05	81.66/1.01	83.76/0.61	85.55/1.02
META+MSA+MIA	-	-	85.21/1.06	83.69/0.66
META+MSA+MIA+ emo. & offens. as i/p features	80.28/0.23	82.40/0.11	85.95/1.03	87.17/0.49

Table 12: Results of the single task stance detection models in varying combinations

the model with the believer stance, but the presence of the insult, toxicity along with anger, disgust, and negative emotions helps identify the correct attitude of the tweet. Further, examples 4 and 5 (Figure 3) suggest that the non-toxic levels along with the trust, anticipation, and positive sentiment provide better insight into the believe stance of the tweet. This inclusion of offending and emotional aspects in our dataset allows the model to use additional information when classifying the tweet's stance.

A.2 Distribution of Emotions and Offensive Content in Benchmark Stance Detection Datasets

In this section, we use the data annotation strategies of emotion recognition and offensive language identification (section 3.1) to identify the distribution of emotions and offensive content in publicly available datasets: ClimateStance-2022 and Semeval2016. Tables 8 and 9 represent the percentage of emotions and offensive expressions for the ClimateStance-2022 dataset. Tables 10 and 11 represent the proportion of emotions and offensive expressions in the "Climate Change is a Real Concern" target of the SemEval-2016 dataset. The distribution for other targets can be found here[10].

B EXPERIMENTS & RESULTS

We experimented with different values for different hyper-parameters, which you can read about here[10]. Table 12 shows the corresponding macro precision and macro recall scores from Table 3. The precision and recall scores of Table 4 can be found here[10].

[10]https://github.com/apoorva-upadhyaya/Emotion_Offensive_Aided_Stance