Sarcasm-based Tweet-level Stress Detection

Prashanth KVTKN[a,*,1], Tene Ramakrishnudu[a,2]

[a]*Department of Computer Science and Engineering, National Institute of Technology, Warangal*

ARTICLE INFO

Keywords:
Psychological stress
Social media
Tweet-level stress detection
Sarcasm
Kernel PCA

Abstract

Psychological stress has evolved as an important health concern across the globe. The vulnerability to stress and the ramifications of it have only worsened during the time of the COVID-19 pandemic. This necessitates a timely diagnosis of stress before the condition progresses to chronicity. In this context, the popularity of social media like Twitter, where large numbers of users share opinions without any social stigma, has emerged as a major resource of human opinions. This has led to an increased research interest in social media-based stress detection techniques. However, tweet-level stress detection techniques in the literature have left a void in leveraging the text information in tweets, especially the presence of sarcastic expressions in the tweet's text content. To this end, a novel method called *"Sarcasm-based Tweet-Level Stress Detection"* (STSD) is proposed in this work with the modification of the logistic loss function to detect tweet-level stress by availing the information of sarcasm that exists in the tweet-content. The principle of the STSD model is to minimise the loss for non-sarcastic tweets while maximising the loss for sarcastic tweets. Furthermore, an extensive preprocessing and dimensionality reduction is performed using *kernel principal component analysis* (kernel PCA) to improve the performance by reducing the dimensions. The experimental results show that the proposed STSD model, when applied along with kernel PCA, records a significant improvement in accuracy by a minimum of 5.25% and a maximum of 9.19% over baseline models. Also, there is an increment in F1-score by at least 0.085 points and a maximum of 0.164 points when compared to the baseline models.

1. Introduction

Chronic stress causes severe physiological problems. Depression, a major cause of suicides across the globe, is due to the persistence of chronic stress over a long period of time (Glaser and Kiecolt-Glaser, 2005). Hence, psychological stress needs to be diagnosed early before it turns chronic. The COVID-19 pandemic has devastated the daily lives of people all over the world, with more than 409 million people infected and over 5.7 million deaths reported globally until January 31, 2022[3]. The pandemic has also changed the lifestyle due to isolation and quarantines, which, along with an unhealthy diet, intensify stress, resulting in an increase in cardiovascular illnesses, especially in women (Mattioli, Sciomer, Maffei and Gallina, 2021). It goes without saying that the pandemic and its ramifications have made people even more vulnerable to psychological stress. Also, the world's second and third most populous countries, India and the United States of America, have been severely affected by the second and third waves of the COVID-19, respectively (Asrani, Eapen, Hassan and Sohal, 2021) [4]. And the impact was observed in social media usage as well as in the population (Chhatwani, Mishra and Rai, 2022; Sv, Lathabhavan and Ittamalla, 2021). Unsurprisingly, people during the period of isolation and quarantine have increased their usage of social media communication with peers for mobilising emotional support—a process called the buffer effect (Marzouki, Aldossari and Veltri, 2021). Moreover, it is observed that due to the ripple effect of COVID-19, the pandemic is acting as a stressor [5], impacting psychological stress on people of various walks of life, ranging from students to migrant workers (Agarwal, 2020).

A variety of studies have been conducted to investigate the economic as well as social consequences of COVID-19 (Zhang, Lyu, Liu, Zhang, Wang, Luo et al., 2021; Baker, Bloom, Davis and Terry, 2020; Nicola, Alsafi, Sohrabi, Kerwan, Al-Jabir, Iosifidis, Agha and Agha, 2020), and have revealed that COVID-19 had a strong impact on the

*Corresponding author
kasturiprashanth@student.nitw.ac.in (P. KVTKN); trk@nitw.ac.in (T. Ramakrishnudu)
ORCID(s):
[1]Research Scholar, Department of Computer Science and Engineering, National Institute of Technology, Warangal, Telangana, India, 506001
[2]Assistant Professor, Department of Computer Science and Engineering, National Institute of Technology, Warangal, Telangana, India, 506001
[3]https://covid19.who.int/
[4]https://www.sciencedaily.com/releases/2022/02/220210154230.htm
[5]Any event that acts as a trigger for stress is a stressor (Lin, Jia, Nie, Shen and Chua, 2016)

mental health of people all over the world. These studies discovered that depression, post-traumatic stress disorder, anxiety, and stress symptoms were reported at elevated rates during COVID-19 than the previous (Xiong, Lipsitz, Nasri, Lui, Gill, Phan, Chen-Li, Iacobucci, Ho, Majeed et al., 2020). Females, young age groups, students, and people with low education levels are especially vulnerable to depression during the pandemic (Xiong et al., 2020). Because of the changes it brought about, the pandemic had a negative impact on people's psychological health. For example, it is noted that people have undergone increased levels of stress due to social isolation during and after the nationwide lockdown (Le, Dang, Toweh, Nguyen, Le, Do, Phan, Nguyen, Pham, Ta et al., 2020). As a result of social isolation and restricted mobility, there has been an increased use of virtual platforms like social media among the public during the period of the COVID-19 pandemic (Huo and Turner, 2019). People are increasingly expressing themselves through social media platforms such as Facebook, Twitter, and Instagram. Therefore, social media emerged as a valuable resource for gathering information about public opinion and their mental health conditions (Huo and Turner, 2019).

With the greater reach of social media and substantial change in lifestyle owing to measures undertaken as part of the COVID-19 protocol, there is an increasing trend of people sharing opinions related to personal life in tweets to express emotions, especially those of negative emotions (Guntuku, Sherman, Stokes, Agarwal, Seltzer, Merchant and Ungar, 2020). The surge in negative emotions reflected in tweets related to COVID-19 gives an opportunity to detect automatic tweet-level stress. It is also understood that text features have the greatest individual contribution to stress detection (KVTKN and Ramakrishnudu, 2021; Lin, Jia, Qiu, Zhang, Shen, Xie, Tang, Feng and Chua, 2017). But the size of tweets acts as a constraint. To utilise the information present in the verbal part of the tweets, various solutions were proposed in some of the earlier works (Coppersmith, Harman and Dredze, 2014; De Choudhury, Gamon, Counts and Horvitz, 2013b; De Choudhury, Counts and Horvitz, 2013a; Lin et al., 2017; Lin, Jia, Guo, Xue, Huang, Cai and Feng, 2014a; Lin, Jia, Guo, Xue, Li, Huang, Cai and Feng, 2014b; Xue, Li, Jin, Feng, Clifton and Clifford, 2014; Xue, Li, Zhao, Jia, Feng, Yu and Clifton, 2016). The majority of these works rely on manual annotation of tweets, which is time-consuming and prone to error. While in (KVTKN and Ramakrishnudu, 2021), a new feature, *Sarcasm_Level*, has been developed apart from utilising a set of previous tweets to overcome the problem of sparsity in data for the detection of tweet-level stress.

Moreover, the utility of sarcasm, which exists in the tweet's text data, in detecting stress is yet to be explored to its full potential. Sarcasm, a word play that hides the original intent of the speaker, has an important role in conveying the opinion of users and affects their mood (Frenda, Cignarella, Basile, Bosco, Patti and Rosso, 2022). Sarcasm is subjective and is relatively tough to detect as compared to sentiment. In much of the existing work related to stress detection, the role of sarcasm needs to be investigated thoroughly (Coppersmith et al., 2014; De Choudhury et al., 2013b,a; Lin et al., 2017, 2014a,b; Xue et al., 2014, 2016). The main intention of its usage is to express contradictory emotion than what is explicitly specified. In addition, due to the implicit presence of sarcasm within the tweet's text content, the influence of sarcasm in stress detection helps in maximising the utilisation of information from the text. In this regard, in (KVTKN and Ramakrishnudu, 2021), the authors employed a new attribute whose computation is derived from the principle of illocutionary sarcasm in order to capture sarcasm to improve the performance of stress detection.

To improve the performance of tweet-level stress detection by utilising the sarcastic information present in the tweet, in this work, a novel approach called Sarcasm-based Tweet-level Stress Detection (STSD) has been developed. To this end, preprocessing of tweets is performed to clean up the noise in the data. Later, the data collected for the problem is organised into two sets based on the sarcasm value computed as described in section 3.3.3. Then the proposed model is trained to predict stress at tweet-level. The proposed STSD model performs better than baseline machine learning(ML) models as it incorporates additional information from the tweet's content through the value of sarcasm. For more generalizability, the proposed model is evaluated using two different validation datasets, collected before and during the third wave of COVID-19 in India. Furthermore, dimensionality reduction techniques are employed to increase the overall performance of the proposed STSD model. To the best of our knowledge, this is the first work in tweet-level stress detection that develops a model to detect stress by utilising sarcastic content in the tweets related to the COVID-19 pandemic. This is also the first work to incorporate dimensionality reduction techniques for the solution of stress detection at tweet-level. The contributions from this work are presented below:

- Developing a model called Sarcasm-based Tweet-level Stress Detection (STSD) to detect stress based on sarcastic content present in the tweets by modifying the loss function of logistic regression.
- An extensive data preprocessing analysis to reduce the number of dimensions and to improve the performance of the proposed STSD model.

- Implementation of a tweet-level stress detection technique on the COVID-19 tweets' data, collected during the second and third waves of the pandemic in India, by utilising information on sarcasm.

The rest of the paper is organised as follows. Section 2 covers the background literature. The problem statement and the proposed model are presented in Section 3. The experimental setup and data description are given in Section 4. The results and discussion are presented in Section 5. Conclusion and future research are discussed in Section 6.

2. Background

Automatic stress detection from social media is of increasing interest in the research community. But the main drawback of these works is manual labeling and lack of inclusion of sarcasm as a clue to detect the stress. In (Thelwall, 2017), an intelligent system called *TensiStrength* was developed in order to automatically detect the label of a stress class from a large collection of short-length text messages, and this information was later utilised in intelligent and smart transport systems. Nevertheless, it did not consider the utilisation of sarcasm present in text content to detect stress levels from a given message.

In the last few years, deep learning-based solutions for social media-based stress detection have secured the interest of the research community. In (Lin et al., 2014b), one of the first works to use deep-learning-based solutions to this problem, various hand-crafted features are extracted at both the tweet and user levels. These features are then loaded into a convolutional neural network, which generates user-level attributes that are modality invariant. Following that, a deep neural network-based model was developed for detecting stress (Lin et al., 2014b). In (Lin et al., 2016), a multi-task learning method is employed to detect stress as well as to determine the subjects of the stressor and its events from a designated post. Moreover, (Lin et al., 2016) employs convolutional neural networks to extract features at both the tweet-level and word-level. However, because the method depends on manual labelling, it cannot scale to larger datasets. The authors of (Lin et al., 2017) developed a model that is a conglomerate of convolutional neural networks and partially labelled factor graphs to determine stress at user-level. Furthermore, (Lin et al., 2017) conducted a methodical scrutiny on the interrelationship between users' stress states and social interaction networks, concluding that users with stress have an excess of sparse connections by 14% in comparison to users with no stress. In (Wang, Wang, Li, Zhang and Wang, 2020), a fusion net model was developed using multi-task deep learning on text vectors for detecting depression in the users. The data is labelled on the criteria of the existence of tweets with phrase patterns such as “I feel stressed” or “I feel stressed this week” in the literature (Lin et al., 2017, 2014b, 2016), as it has been proven that the phrase structures based on “I feel” assist in better emotional analysis (Kamvar and Harris, 2011). But none of these methods, for social media-based stress detection, have incorporated information related to sarcasm into their mechanisms.

Recently, a deep learning-based solution has been investigated in (Mundotiya and Yadav, 2021) for identifying clickbaits in social media posts, which can mislead users while also regulating their stress levels. In (Gandhi, Kumar, Babu and Karthick, 2021), a methodology built upon convolutional neural networks (CNN) and long-term short-term memory (LSTM) is proposed. It detects a tweet's polarity by extracting a multitude of features to represent the context. In (Martins, Almeida, Henriques and Novais, 2021), a solution for identifying depression clues in text has been developed based on a procedure for detecting Twitter profiles that are depressive. It incorporates techniques from various fields, such as sentiment analysis, machine learning, and natural language processing. But, none of these tools capture the information related to sarcasm present in tweet content to detect stress.

Sarcasm is a type of communication wherein the meaning conveyed is the opposite to the meaning stated explicitly (Liu, Chen, Ou, Wang, Yang and Lei, 2014). In (Rajadesingan, Zafarani and Liu, 2015), a behavioural modelling framework is employed to detect sarcasm in users' tweets. It is proposed in (Mukherjee and Bala, 2017) that for the effective automatic recognition of sarcasm in Twitter data, both text content features and authorial style criteria are required. In (Sundararajan and Palanisamy, 2020), a fuzzy-logic-based approach is presented for detecting sarcasm in tweets and classifying sarcasm into four categories based on the intensity of hurt in the statement. In addition, various forms of expressing sarcasm are described in (Camp, 2012; Joshi, Bhattacharyya and Carman, 2017; KVTKN and Ramakrishnudu, 2021).

In (KVTKN and Ramakrishnudu, 2021), the authors have proposed two solutions for overcoming the problem of sparsity in the detection of stress at the granularity of tweets. First, an attribute called *Sarcasm_Level* is extracted to reflect the sarcasm that exists in the content of a tweet for the detection of tweet-level stress. Second, a novel method called Neighborhood-based Tweet-level Stress Detection (NTSD) is proposed to utilise the information from neighbouring tweets to increase the availability of related data. And it was concluded that the use of sarcasm as a

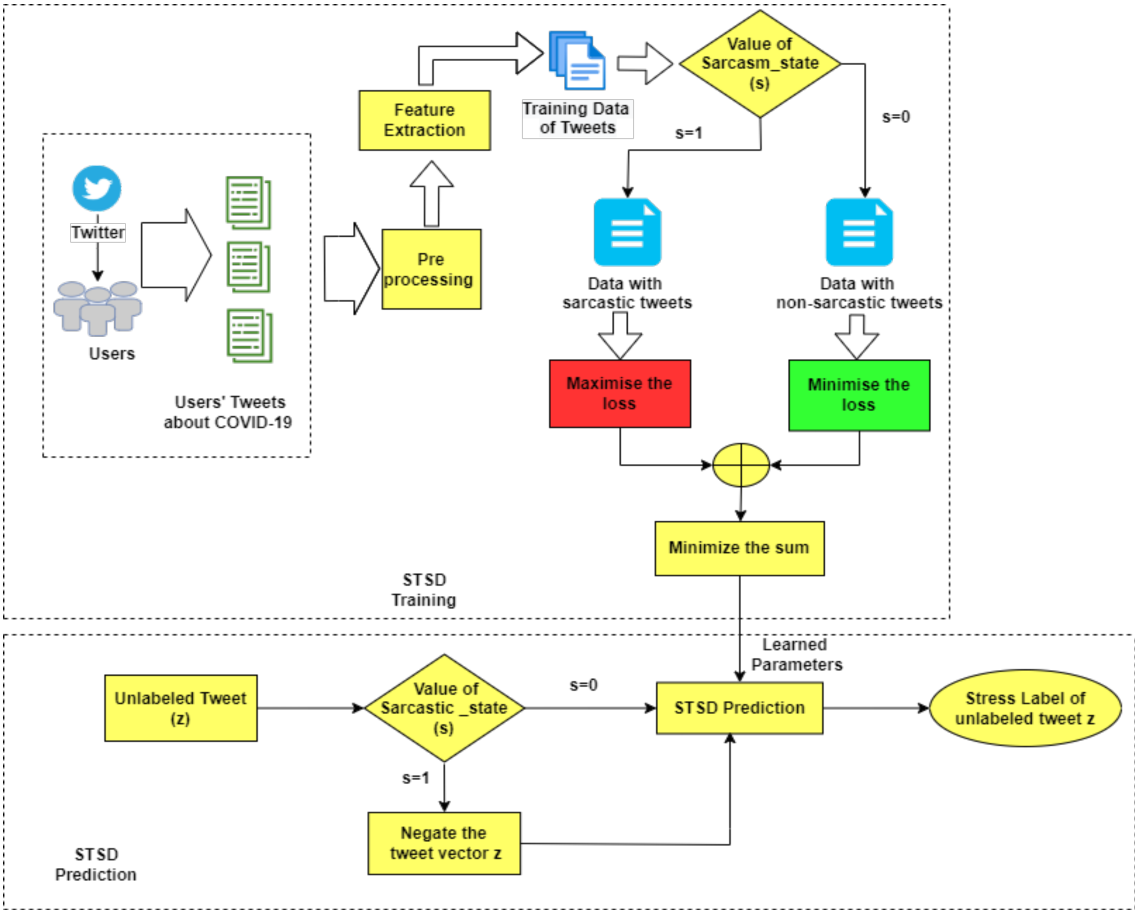


Figure 1: Framework of the STSD model

binary-valued feature helps in increasing the performance of stress detection. Also, the NTSD in combination with the new attribute outperforms the baseline models. But the model is computationally intensive as it includes the content of neighbourhood tweets as well.

Furthermore, dimensionality reduction analysis for improving the performance of tweet-level stress detection is not addressed in the literature (KVTKN and Ramakrishnudu, 2021; Lin et al., 2017, 2016; Xue et al., 2016; Zhao, Jia and Feng, 2015; Pratama and Sarno, 2015; Lin et al., 2014a). It is observed that Kernel Principal Component Analysis (PCA) helps in obtaining the dimensionality reduction when the classes are linearly non-separable (Satour, Benyacoub, El Mahrad and Kacimi, 2021).

In this paper, a new logistic regression-based approach to detect tweet-level stress from COVID-19 tweets by explicitly using information about sarcasm is proposed. The information about sarcasm is computed from a variable called *Sarcasm_state* proposed in this work. By incorporating the concept of sarcasm, this technique is designed to maximise the utility of textual information within the tweet. In addition, this work also utilises the dimensionality reduction technique, which was not addressed in earlier literature on tweet-level stress detection. In this work, a non-linear PCA technique called polynomial kernel PCA is applied to enhance the performance of the proposed model.

3. Methodology

In this Section, the concept, formulation, and procedure of the proposed model of Sarcasm-based Tweet-level Stress Detection (STSD) are discussed. The aim of this approach is to develop a classification model to detect stress

Table 1
Description of the symbols employed in this work

Symbol	Description
U	Collection of users.
u	Any particular user u, where $u \in U$.
$\mathbf{x_i}$	It is feature vector of i-th tweet in dataset D and there exists some owner $u \in U$, for this tweet.
N	The cardinality or the number of tweets present in the dataset D.
M	Count of attributes in the model or length of the tweet's feature vector $\mathbf{x_i}$.
s_i	The value of *Sarcasm_state* for the tweet $\mathbf{x_i}$.
D_s	The set of sarcastic tweets in the training data.
D'_s	The set of non-sarcastic tweets in the training data, $D = D_s \bigcup D_s'$.
$y_i \in \{0, 1\}$	Class label denoting Stress state of the tweet $\mathbf{x_i} \in D$.
Y	Collection of class labels of all tweets, $\mathbf{x_i}$, $\forall i \in 1, 2, ..., N$, where $N = \|D\|$.
$\mathbf{w}$	The weight vector of features. It is of size M, $\mathbf{w} = [\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_M]^T$.
$p(\mathbf{x})$	The sigmoid function or logistic function, defined as $p(\mathbf{x}) = \frac{1}{1+\exp(-\mathbf{w}.\mathbf{x})}$.
$f(\mathbf{w})$	The loss function for the proposed model of STSD.
∇f	Gradient for the loss function, $f(\mathbf{w})$ with respect to parameter vector $\mathbf{w}^T$.
∇f_i	Gradient for loss function at ith sample of training data, D, with respect to weight vector $\mathbf{w}^T$.
η	Learning rate.
$P_s(y_i\|\mathbf{x_i})$	Prediction probability for STSD model, specifying the probability that the class y_i corresponds to tweet $\mathbf{x_i}$.

at tweet-level by availing sarcasm information, thereby increasing the utilisation of the text information for better performance.

3.1. Problem Concept

The framework representing the concept of the proposed STSD model is depicted in figure 1. The figure shows the proposed Sarcasm-based Tweet-level Stress Detection (STSD) method consists of two phases - STSD training and STSD prediction. During training, the tweets are preprocessed and the feature extraction is performed. The resulting processed training data is categorised into two sets based on the value of *Sarcasm_state*. In this approach, the label for stress is learnt by minimising the loss for non-sarcastic tweets and maximising the loss for sarcastic tweets, as sarcastic tweets generally have subtle sentiment that is contrary to the explicit emotion specified. For this purpose, the value of sarcasm is utilised in the loss function of the proposed approach. During the prediction stage, the unlabelled input vectors are processed based on the value of the *Sarcasm_state*. The symbols employed to formulate and implement this problem are described in Table 1. The detailed procedure of all the steps involved in the process is presented in section 3.3.

3.2. Problem Statement

The problem is to find a function G that takes an unlabeled tweet, z, and considers its *Sarcasm_state*, s, to produce a label, y, for the tweet, where $y \in \{0, 1\}$ is class label denoting stress of the tweet. The parameters of the function G are learned from training data such that the log-likelihood loss for sarcastic tweets ($s_i = 1$) is maximised while the log-likelihood loss for non-sarcastic tweets ($s_i = 0$) is minimized. The function G is described as $G : D \rightarrow Y$, where Y is the set of class labels. Subsequently, the classifier G is employed to predict the label y of the unlabeled tweet $\mathbf{z}$ by utilizing the information of its *Sarcasm_state*.

3.3. Proposed Method : Sarcasm-based Tweet-level Stress Detection (STSD)

In this section, the detailed process of the proposed STSD method is described, from the initial preprocessing of tweets to prediction. Also, the idea and working mechanism of the proposed Sarcasm-based Tweet-level Stress Detection (STSD) are presented. Finally, dimensionality reduction techniques for improving the performance of the proposed STSD model are also discussed.

3.3.1. Preprocessing

After the collection of tweets, preprocessing should be performed to remove the noise and missing information in the tweets. Initially, the data is cleaned by removing the tweets that have no textual content in them. Also, tweets that contain only URLs are removed. In addition, all tweets that contain text in a language other than English are removed.

This filtered set of tweets is used for extracting features relevant for the classification of tweet-level stress. Later, the filtered set of tweets is used in exploratory data analysis.

3.3.2. Feature Extraction

In order to predict tweet-level stress more accurately, the information from a tweet's text content and clues of sarcasm in it are leveraged. For this purpose, linguistic-content features are extracted similar to earlier works in the literature (KVTKN and Ramakrishnudu, 2021; Lin et al., 2017). The social attributes related to a tweet's social engagement computed in the problem are described as follows:

- Attributes of text-content
  - *Count Vector of Categories*: Library of EMPATH [6] is employed to obtain the word count associated with ten categories of linguistic psychological inferences for the tweet's text content. Family, health, school, academics, exams, disease, business, interpersonal, office, and informal language are among the categories. This vector is of length ten.
  - *Verb Vector*: This is a two-element vector storing the quantities of verbs in a tweet, for both positive and negative polarities.
  - *Adjective Vector*: This is a two-dimensional vector storing the quantities of adjectives in a tweet, for both positive and negative polarities.
  - *Degree adverbs Vector*: This is a two-element vector storing the strength of adverbs in a tweet, for both positive and negative polarities. The strength of the verb is computed as specified in the literature (Lin et al., 2017; KVTKN and Ramakrishnudu, 2021; Lin et al., 2014a,b). The degree scales in the range of $[-3, 3]$. "*I feel stressed*" has a degree of -1, whereas "*I feel more stressed*" takes degree -2. Similarly, ""*I am joyful*" takes degree $+1$, whereas "*I am extremely joyful*" takes the degree $+3$.
  - *Emoji Vector*: This is a two-element vector storing the quantities of emojis and emoticons in a tweet, for both positive and negative polarities. Python's NLTK (Natural Language Tool Kit) and Emoji libraries are used to extract emoji sentiment.
  - *Punctuation Vector*: It is a 3 dimensional vector. It stores the number of punctuation marks in a tweet, such as exclamation marks (!), question marks (?), and dots (...). The punctuation marks reflect the emotion of the tweet (KVTKN and Ramakrishnudu, 2021).
- Social attributes vector
  - This feature vector signifies the tweet's social engagement. It is a vector of three elements and stores for a given tweet the number of favourites or likes, the number of retweets, and the number of comments.

3.3.3. Computing Sarcasm_State value

Based on the work (KVTKN and Ramakrishnudu, 2021), a new value called *Sarcasm_state* is derived to represent the sarcasm that exists in the tweet's text, built on the notion of illocutionary sarcasm. In this work, the computation of sarcasm is further modified to capture the contradictory emotions within words and hash tags apart from reflecting the inconsistency between the polarity of text content (like words and hashtags) and the polarity of non-verbal expressions like emojis and emoticons. Broadly, the *Sarcasm_state* is allocated a value of 1, when there is a disagreement between the polarity of the majority of words and the polarity of the majority of the hashtags within the text part of the tweet. In addition, the *Sarcasm_state* is given a value of 1 when there is a disagreement between the polarity of the majority of words in a tweet's text content and the polarity of the majority of non-verbal expressions like emojis and emoticons. Similarly, the *Sarcasm_state* is assigned a value of 1 in cases where there is a disagreement between the polarity of the majority of hashtags and the polarity of the majority of non-verbal expressions like emojis and emoticons. In all the remaining cases, where there is similar polarity for the majority of text content and the majority of non-verbal expressions like emojis and emoticons, the *Sarcasm_state* is assigned a value of 0.

[6] EMPATH is an open source library for obtaining words counts based on psychology and linguistics (Fast, Chen and Bernstein, 2016).

Algorithm 1: Computation of *Sarcasm_state*

input : $\mathbf{x}$, a tweet in the set D, the labeled Training Data set of tweets.
output: Sarcasm_state, $\mathbf{s}$

pos ←number of words in the tweet $\mathbf{x}$ with value of polarity being positive
neg ←number of words in the tweet $\mathbf{x}$ with value of polarity being negative
e_{pos} ←number of positive emojis and emoticons in the tweet $\mathbf{x}$
e_{neg} ←number of negative emojis and emoticons in the tweet $\mathbf{x}$
h_{pos} ←number of hashtags with positive polarity
h_{neg} ←number of hashtags with negative polarity in the tweet $\mathbf{x}$
if $e_{pos} > e_{neg}$ & $pos < neg$ **then**
    $Sarcasm\_state \leftarrow 1$
end
if $e_{pos} < e_{neg}$ & $pos > neg$ **then**
    $Sarcasm\_state \leftarrow 1$
end
if $h_{pos} > h_{neg}$ & $pos < neg$ **then**
    $Sarcasm\_state \leftarrow 1$
end
if $h_{pos} < h_{neg}$ & $pos > neg$ **then**
    $Sarcasm\_state \leftarrow 1$
end
if $h_{pos} > h_{neg}$ & $e_{pos} < e_{neg}$ **then**
    $Sarcasm\_state \leftarrow 1$
end
if $h_{pos} < h_{neg}$ & $e_{pos} > e_{neg}$ **then**
    $Sarcasm\_state \leftarrow 1$
else
    $Sarcasm\_state \leftarrow 0$
end

3.3.4. Training of the proposed STSD model

In this section, the training procedure and related algorithms of the proposed STSD model are discussed. To understand the loss function and training procedure of the proposed model, the idea and working mechanism of the proposed model of STSD are presented.

The principle of the proposed STSD model is to maximise the likelihood of a tweet belonging to the class of stress if it is a non-sarcastic tweet and to minimise the likelihood of a tweet belonging to the class of stress if it is a sarcastic tweet. This can be interpreted as-to minimise the loss for non-sarcastic tweets and maximise the loss for sarcastic tweets. From the work (James, Witten, Hastie and Tibshirani, 2013), it is noticed that, in a logistic regression model, for any feature vector $\mathbf{x}$, whose corresponding class label is y, the likelihood of the tweet $\mathbf{x}$ belonging to the class y is given as:

$$p(\mathbf{x})^y(1 - p(\mathbf{x}))^{(1-y)} \tag{1}$$

Where, $p(\mathbf{x})$ is sigmoid or logistic function, defined as follows:

$$p(\mathbf{x}) = \frac{1}{1 + \exp(-\mathbf{w}.\mathbf{x})} \tag{2}$$

Therefore, the log-likelihood in this scenario is given as:

$$y\log(p(\mathbf{x})) + (1 - y)\log(1 - p(\mathbf{x}))$$

According to the principle of the proposed model, the likelihood of the model is constructed based on the likelihoods of sarcastic and non-sarcastic tweets as follows:

1. For non-sarcastic tweets ($s_i = 0$), the likelihood of belonging to the class stress is maximized. Hence, the likelihood of the tweets $\mathbf{x_i} \in {D_s}'$ is

$$p(\mathbf{x_i})^{y_i}(1 - p(\mathbf{x_i}))^{1-y_i}$$

Then the log-likelihood in this scenario is given by

$$y_i \log(p(\mathbf{x_i})) + (1 - y_i)\log(1 - p(\mathbf{x_i})) \tag{3}$$

2. The likelihood of belonging to the class stress is to be minimised for sarcastic tweets ($s_i = 1$). This is achieved by maximising the likelihood of belonging to the class non-stressed and minimising the likelihood of belonging to the class stress. Hence, the likelihood for the tweets $\mathbf{x_i} \in D_s$ is

$$p(\mathbf{x_i})^{1-y_i}(1 - p(\mathbf{x_i}))^{y_i}$$

Then the log-likelihood in this case is given as

$$(1 - y_i)\log(p(\mathbf{x_i})) + y_i \log(1 - p(\mathbf{x_i})) \tag{4}$$

Hence, using the equations (3) and (4), the total log-likelihood, $l(\mathbf{w})$, of the proposed model, computed for all the tweets in the training dataset, is given as:

$$l(\mathbf{w}) = \sum_{i=1}^{N}(1 - s_i)\Big\{ y_i \log(p(\mathbf{x_i})) + (1 - y_i)\log(1 - p(\mathbf{x_i}))\Big\} + \sum_{i=1}^{N} s_i\Big\{ y_i \log(1 - p(\mathbf{x_i})) + (1 - y_i)\log(p(\mathbf{x_i}))\Big\}$$

Then the loss of the proposed model STSD, $f(\mathbf{w})$, is computed as negative value of log-likelihood $l(\mathbf{w})$:

$$f(\mathbf{w}) = -l(\mathbf{w}) = -\Big(\sum_{i=1}^{N}(1-s_i)\Big\{ y_i \log(p(\mathbf{x_i}))+(1-y_i)\log(1-p(\mathbf{x_i}))\Big\}+\sum_{i=1}^{N} s_i\Big\{ y_i \log(1-p(\mathbf{x_i}))+(1-y_i)\log(p(\mathbf{x_i}))\Big\}\Big) \tag{5}$$

In other words, the concept of the proposed STSD model is defined as minimising the loss for non-sarcastic tweets while maximising the loss for sarcastic tweets. For solving the parameters of the proposed model, gradient descent based algorithms are implemented. The optimization algorithms of gradient descent require the computation of the gradient of the loss function of the proposed model, which is derived in Equation 6. As the prediction function, p, is a composite function of both weight vector, $\mathbf{w}$ and feature vector $\mathbf{x_i}$, the computation of the gradient of loss function of proposed STSD model is as follows:

$$\frac{\partial f}{\partial \mathbf{w}} = \nabla f$$

$$= -\Bigg[\sum_{i=1}^{N} \frac{\partial\Big\{(1 - s_i)\Big\{ y_i \log(p(\mathbf{x_i})) + (1 - y_i)\log(1 - p(\mathbf{x_i}))\Big\}\Big\}}{\partial \mathbf{w}} + \sum_{i=1}^{N} \frac{\partial\Big\{(s_i)\Big\{ y_i \log(1 - p(\mathbf{x_i})) + (1 - y_i)\log(p(\mathbf{x_i}))\Big\}\Big\}}{\partial \mathbf{w}}\Bigg]$$

$$= -\Bigg[\sum_{i=1}^{N}(1-s_i)\Big\{y_i\frac{1}{p(\mathbf{x_i})}\frac{\partial p(\mathbf{x_i})}{\partial \mathbf{w}} + (1-y_i)\frac{1}{(1-p(\mathbf{x_i}))}\frac{\partial(1-p(\mathbf{x_i}))}{\partial \mathbf{w}}\Big\} + \sum_{i=1}^{N}s_i\Big\{y_i\frac{1}{(1-p(\mathbf{x_i}))}\frac{\partial(1-p(\mathbf{x_i}))}{\partial \mathbf{w}} + (1-y_i)\frac{1}{p(\mathbf{x_i})}\frac{\partial p(\mathbf{x_i})}{\partial \mathbf{w}}\Big\}\Bigg]$$

$$= -\Bigg[\sum_{i=1}^{N}\Big\{(1-s_i)(y_i - p(\mathbf{x_i}))x_{im}\Big\} + \sum_{i=1}^{N}\Big\{s_i(1-y_i-p(\mathbf{x_i}))x_{im}\Big\}\Bigg]$$

$$\implies \nabla f = -\Big[\sum_{i=1}^{N}\Big\{y_i - 2y_is_i + s_i - p(\mathbf{x_i})\Big\}x_{im}\Big] \tag{6}$$

The whole process describing the training and prediction of the proposed STSD model is presented in the Algorithm 2. The Algorithm describes how the information of sarcasm present in the tweet is extracted from each tweet of the training data and is utilised for learning the parameters using gradient descent. This can be seen from steps 4-9 in Algorithm 2. The procedure call at step 6 invokes the Algorithm 1 and returns the value of sarcasm. The procedure call of *Gradient_Descent_STSD* at step 9 invokes Algorithm 3, which returns learnt parameters, **w**. To predict labels for unlabeled tweets, first the value of *Sarcasm_state* of the tweets is determined and then the label is predicted with the help of sigmoid function.

Theorem 1. *The loss function of the proposed STSD model is a special form of logistic regression loss.*

Proof. From the loss function of the proposed STSD model, presented in Equation (5), it is observed that at any given iteration, only one component of the loss function is computed due to the binary variable s_i. For the tweets with $s_i = 1$, only the second component is computed, while only the first component is computed for the tweets with s_i=0.
From the definition of the sigmoid function (Han, Pei and Kamber, 2011):

$$p(-\mathbf{x}) = \frac{1}{1+\exp(-\mathbf{w}.(-\mathbf{x}))} = \frac{\exp(-\mathbf{w}.\mathbf{x})}{1+\exp(-\mathbf{w}.\mathbf{x})} = 1 - \frac{1}{1+\exp(-\mathbf{w}.\mathbf{x})} = 1 - p(\mathbf{x})$$

Hence, using the above relation, the loss function is rewritten as:

$$f(\mathbf{w}) = -l(\mathbf{w})$$
$$= -\Big(\sum_{i=1}^{N}(1-s_i)\Big\{y_i\log(p(\mathbf{x_i})) + (1-y_i)\log(1-p(\mathbf{x_i}))\Big\} + \sum_{i=1}^{N}s_i\Big\{y_i\log(p(-\mathbf{x_i})) + (1-y_i)\log(1-p(-\mathbf{x_i}))\Big\}\Big) \tag{7}$$

From Equation (7), it is noticed that the two components in the summation are characterized by the value of *Sarcasm_state*, s_i. The first component of the loss exists only if the value of *Sarcasm_state* vanishes if $s_i = 0$ and the second component exists if $s_i = 1$. Hence, Equation (7) can be rewritten into a single component as follows:

$$f(\mathbf{w}) = -\Big(\sum_{i=1}^{N}\Big\{y_i\log(p(\mathbf{x_{ti}})) + (1-y_i)\log(1-p(\mathbf{x_{ti}}))\Big\}\Big) \tag{8}$$

where,
$\mathbf{x_{ti}} = \mathbf{x_i}$,if $s_i = 0$
$\mathbf{x_{ti}} = -\mathbf{x_i}$, if $s_i = 1$
The Equation (8) is similar to a logistic loss function (James et al., 2013). Hence, it is demonstrated that the proposed STSD model's loss function reduces to a special case of the logistic regression loss. □

Algorithm 2: Sarcasm-based tweet-level stress detection

Input: Labeled Dataset of Tweets, D

Output: Model object with learnt parameters and labels for unlabeled tweets

Function STSD(D)**:**
- $Y \leftarrow \left\{ y_i \middle| (\mathbf{x_i}, y_i) \in D, \quad 1 \leq i \leq N \right\}$
- Perform initial preprocessing of the data
- $\mathbf{s} \leftarrow \phi$
- **for** $x_i \in D$ **do**
  - $s\_i \leftarrow Sarcasm\_state(\mathbf{x_i})$;
  - $\mathbf{s} \leftarrow \mathbf{s} \bigcup s\_i$
- **end**
- $\mathbf{w} \leftarrow Gradient\_Descent\_STSD(D, \mathbf{s})$
- $p(\mathbf{x}) \leftarrow \frac{1}{1+e^{-\mathbf{w}.\mathbf{x}}}, \quad \forall$ tweet $\mathbf{x}$
- $\mathbf{Z} \leftarrow Unlabeled\_tweets$
- $Y_Z \leftarrow \phi$
- **for** $z \in Z$ **do**
  - $s_z \leftarrow Sarcasm\_state(\mathbf{z})$;
  - **if** $s_z = 1$ **then**
    - $prediction\_probability \leftarrow p(-\mathbf{z})$
  - **else**
    - $prediction\_probability \leftarrow p(\mathbf{z})$
  - **end**
  - **if** $prediction\_probability > 0.5$ **then**
    - $y_z \leftarrow 1$
  - **else**
    - $y_z \leftarrow 0$
  - **end**
  - $Y_Z \leftarrow Y_Z \bigcup y_z$
- **end**
- $result\_object \leftarrow \{\mathbf{w}, Y_Z\}$

return $result\_object$

The gradient descent and stochastic gradient descent algorithms for learning the parameters of the proposed STSD model are described in Algorithms 3 and 4, respectively. Both the Algorithms use the gradient of the loss function of the proposed model, ∇f, as specified in Equation (6).

The model's parameters, which have been learnt during training, are used in the prediction function to predict the label for the unlabeled tweets. Based on the principle of STSD, the prediction function for the proposed STSD model is defined with respect to the value of *Sarcasm_state*, s_i, of the tweet $\mathbf{x_i}$ as shown below:

$$P_s(y_i|\mathbf{x_i}) = \begin{cases} P(y_i|\mathbf{x_i}) & \text{, if } s_i = 0 \\ P(y_i|-\mathbf{x_i}) & \text{, otherwise} \end{cases}$$

where, $P(y_i|\mathbf{x_i})$ is sigmoid function and hence, $P(y_i|-\mathbf{x_i}) = p(-\mathbf{x_i}) = \frac{1}{1+\exp(-\mathbf{w}^T.\mathbf{x})}$.

3.3.5. Dimensionality Reduction

In this work, various techniques of dimensionality reduction are applied so as to improve the classification performance of the model. Given the large number of features and sparse nature of the training data of tweet-level stress detection, applying an effective dimensionality reduction technique is essential to reduce the features and to improve the performance of the model (Han et al., 2011). An exploratory analysis of the datasets is conducted by applying

Algorithm 3: Gradient Descent Algorithm for training of STSD Model

Input: Training Data set consisting of D containing the features of tweets in data set, with $|D| = N$ and $\mathbf{s} = \{s_i | i = 1, 2, ..., N\}$, a vector representing the *Sarcasm_state* of each tweet and η, the rate of convergence

Output: Parameter Vector $\mathbf{w}$ of the objective function

Function `STSD_Gradient_Descent(`$D, \mathbf{s}$`):`
  Initialize parameter vector $\mathbf{w}$ to small random values ;
  while *convergence is not reached* **do**
    $\nabla f = \sum_{i=1}^{N}(y_i - 2y_i s_i + s_i - p(\mathbf{x_i}))x_{im}$;
    Update the weight vector, $\mathbf{w} \leftarrow \mathbf{w} + \eta * \nabla f$;
  end
return w

Algorithm 4: Stochastic Gradient Descent Algorithm for training of STSD Model

Input: Training Data set consisting of D containing the features of tweets in data set, with $|D| = N$ and $\mathbf{s} = \{s_i | i = 1, 2, ..., N\}$, a vector representing the *Sarcasm_state* of each tweet and η, the rate of convergence

Output: Parameter Vector $\mathbf{w}$ of the objective function

Function `STSD_Stochastic_Learning(`$D, \mathbf{s}$`):`
  Initialize parameter vector $\mathbf{w}$ to small random values ;
  while *Convergence is not reached* **do**
    for *each training example* $(\boldsymbol{x_i}, y_i)$ **do**
      $\nabla f_i = (y_i - 2y_i s_i + s_i - p(\mathbf{x_i}))x_{im}$;
      Update the weight vector as, $\mathbf{w} \leftarrow \mathbf{w} + \eta * \nabla f_i$
    end
  end
return w

popular dimensionality reduction techniques such as Linear Principal component Analysis (PCA) and Nonlinear or Kernel Principal component Analysis (NPCA) (Lee and Verleysen, 2007).

In linear PCA, the data is projected onto a linear subspace of lower dimensions (Han et al., 2011). But linear PCA does not perform well when the data is linearly non-separable (Satour et al., 2021).

In non-linear or kernel-PCA, the features are transformed into higher dimensions using a non-linear mapping and, later, linear PCA is applied in the transformed higher space (Lee and Verleysen, 2007). Given the high computational cost in computing the transformation to higher dimensions, kernel trick is used. The kernel trick is helpful as it avoids computation of mapping to higher dimensions but allows computation of linear sub-spaces of higher dimensions using kernel functions within the original input space.

The three popular kernels used in kernel PCA are the Linear kernel, Radial Basis Function (RBF) kernel, and Polynomial kernel. In this work, Linear kernel PCA, RBF kernel PCA, and Polynomial kernel PCA are applied on 4 different datasets, and it is observed that polynomial kernel PCA (with degree 3) has good separation of classes when projected onto the two principal components. Hence, the proposed work employs kernel PCA with a polynomial kernel for dimensionality reduction. The detailed discussion of the obtained results using three PCA techniques on all the datasets is presented in section 5.

4. Experimental Setup

This section describes the experimental setup, utilised datasets, and the employed baseline models, which are used to assess the proposed model's performance. All the experiments are conducted using Python with the IDE Jupyter Notebook 4.1.1.

Table 2
The Details of the Datasets

Dataset	Number of tweets	Number of users	COVID-19 Period
D1	7289	1732	Second Wave
D2	16,532	5119	Second Wave
D3	4062	1888	Third Wave
D4	7209	2558	Third Wave

4.1. Dataset Description

To evaluate the model, four datasets of tweets are extracted using Tweepy, Twitter's API. The datasets D1 and D2 are collected during the starting stage of the second wave of COVID-19 in India during the months of February and March, 2021. The rest of the datasets are collected in two stages. Dataset D3 is extracted during the starting stage of the third wave in India, from December 2021 to January 2022. Whereas, dataset D4 is collected during during the peak stage to tail-end of third wave in India -January 2022 to February 2022. The datasets belonging to different time periods are used for implementation to validate the generalising ability of the model.

Tweets are extracted using Tweepy with two types of queries. The tweets are collected using the query "*I feel Stressed*" along with the search words "*COVID-19, covid*". All the tweets collected using this query and keywords are labelled as stressed ($y_i = 1$). While the tweets collected using the query "*I feel relaxed*", "*I don't feel stressed*" along with the search words "*COVID-19, covid*" are labelled non-stressed ($y_i = 0$). This is because, *I feel* pattern-based extraction has proved to be effective (KVTKN and Ramakrishnudu, 2021). The collected datasets are cleaned to remove noise. The details of the datasets collected are presented in Table 2.

4.2. Baseline Models Employed

Popular ML classifiers such as Logistic Regression (LR), Support Vector Machines (SVM), Random Forest (RF), and Naive Bayes (NB) are used to evaluate the performance of the proposed STSD model. There are two reasons why these baseline models are considered. The first is that they are well-known for their performance in text-data classification (Lin et al., 2014a,b; Zhao et al., 2015; Xue et al., 2014, 2016; Xu, Wang and Ai, 2020; KVTKN and Ramakrishnudu, 2021). Second, these models use various kinds of fundamental mathematical concepts to develop the model and, hence, help in obtaining better generalisation of the results of the proposed model (Han et al., 2011; Boser, Guyon and Vapnik, 1992; Hastie, Tibshirani and Friedman, 2009; Breiman, 2001; Kim, Song, Kim, Lee and Cheon, 2018; Koller, Friedman, Džeroski, Sutton, McCallum, Pfeffer, Abbeel, Wong, Heckerman, Meek et al., 2007; KVTKN and Ramakrishnudu, 2021). The abstract details of the utilised base-line models are given below:

1. **SVM:** Support Vector Machines (SVM) are popular classifiers that operate based on the principle of maximum margin separation of data points from a discriminating hyper-plane (Boser et al., 1992; Han et al., 2011). Because of their high performance, they are utilised in a large range of applications involving classification tasks (Han et al., 2011; Boser et al., 1992; Hastie et al., 2009).

2. **RF:** Random Forest (RF) is an ensemble technique for decision tree classifiers. The majority vote is relied upon for the decision (Breiman, 2001; Hastie et al., 2009). Bagging is a technique used in the forest to enhance the stability and consistency of individual decision tree (Hastie et al., 2009). The ensemble decision is formed by the sum of all individual tree decisions (Breiman, 2001; Hastie et al., 2009).

3. **NB:** Naïve Bayes (NB) is a popular binary classification algorithm built upon the concept of Bayes theorem (Han et al., 2011). It predicts posterior probabilities for each class label for a given unlabeled tuple using Bayes theorem (Han et al., 2011). The tuple is assigned to a class with the highest posterior probability. Naïve Bayes assumes class-conditional independence, where each of the input features is independent of each other (Mukherjee and Bala, 2017; Han et al., 2011).

4. **LR:** Logistic regression (LR) is a popular binary classification method that employs a logistic function to forecast the posterior probability for a class (Koller et al., 2007). Logistic regression is a generalised linear model in which the link function is a sigmoid or logistic function (Kim et al., 2018). The proposed STSD model of this work is developed based on the baseline model of logistic regression.

All of these baseline classifiers, as well as the proposed STSD model, are implemented on all the four datasets -D1, D2, D3 and D4.

4.3. Performance Measures

The popular measures of Accuracy and F1-score are used to compare the effectiveness of the proposed model to other baseline models.

Accuracy: The proportion of appropriately predicted data points in comparison to the whole set of data points used in testing (Han et al., 2011). The predictions are organised into groups as follows:

- **True positives (TP):** The number of test samples of the positive stress class that the model properly classifies.
- **True negatives(TN):** The number of test samples of negative stress class that the model properly classifies.
- **False positives (FP):** The number of test samples of negative stress class that the model erroneously classifies.
- **False negatives (FN):** The number of test samples of the positive stress class that the model erroneously classifies.

The accuracy is then computed as follows:

$$Accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)} \tag{9}$$

F1-Score: The harmonic mean of precision and recall is used to compute the F1-Score. If precision, α, is the proportion of correct predictions to all positive predictions and recall, β, is the proportion of correct predictions to total actual positive samples (Chicco and Jurman, 2020), then F1-score is evaluated as :

$$F1 - Score = \frac{2 \times \alpha \times \beta}{(\alpha + \beta)} \tag{10}$$

On the four datasets-D1, D2, D3 and D4-the following experiments were conducted:

1. All the models considered in this work (baseline models and the proposed STSD model) are built with original features without applying any dimensionality reduction technique.
2. Transforming the data by applying the polynomial kernel PCA.
3. Building all the models considered in this work on data of reduced dimensions.

5. Results and Discussion

This section contains the results and their discussion for all the experiments that are performed as part of this work.

5.1. Preprocessing Results

This section discusses the results of exploratory data analysis after applying PCA techniques as part of the preprocessing task as described in section 3.3.5. In this work, to reduce the dimensionality of the data and to mitigate the low performance of the models, a linear PCA technique and a kernel PCA technique with two different kernels—RBF Kernel and Polynomial Kernel—are employed. In the figures 2, 3, 4 and 5, the plots respectively show the projection of the datasets D1, D2 , D3, and D4 on the two principal components after applying various PCA techniques used in this work.

For the dataset D1, the separation of points into stressed and non-stressed classes is more clear after applying a polynomial kernel, as noted from the figure 2. But after applying Linear PCA and RBF kernel PCA, the classes are not linearly separable. Similarly, for dataset D2, the points of the stressed class are hidden behind the non-stressed points after applying polynomial kernel PCA, as seen from the figure 3. For dataset D3, it is observed from figure 4 that the data-points are linearly separable after applying polynomial kernel PCA. Also, in the case of dataset D4, there is a clustering of two classes in both directions of the axes of principal components, as observed in the figure 5.

There is no clear linear separation of the points in the projection after polynomial kernel PCA in the case of dataset D4. Nonetheless, in all the four datasets, the data points are grouped into nearly linearly separable classes after the application of polynomial kernel PCA. Consequently, this helps in better classification. Hence, the polynomial kernel PCA technique is used for feature reduction in this work before building the classifiers.

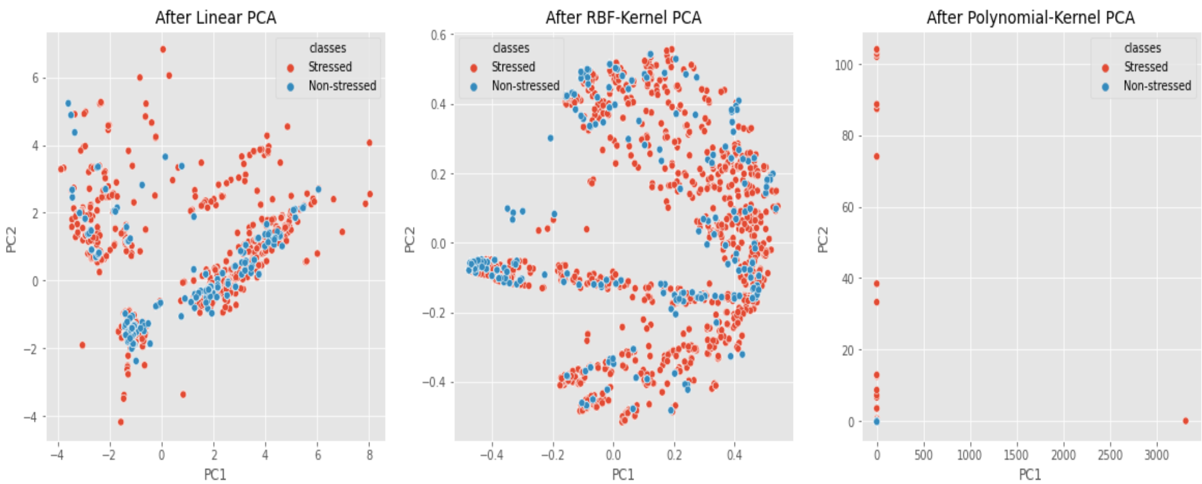


Figure 2: The projection of Dataset D1 on two principal components after different PCA techniques

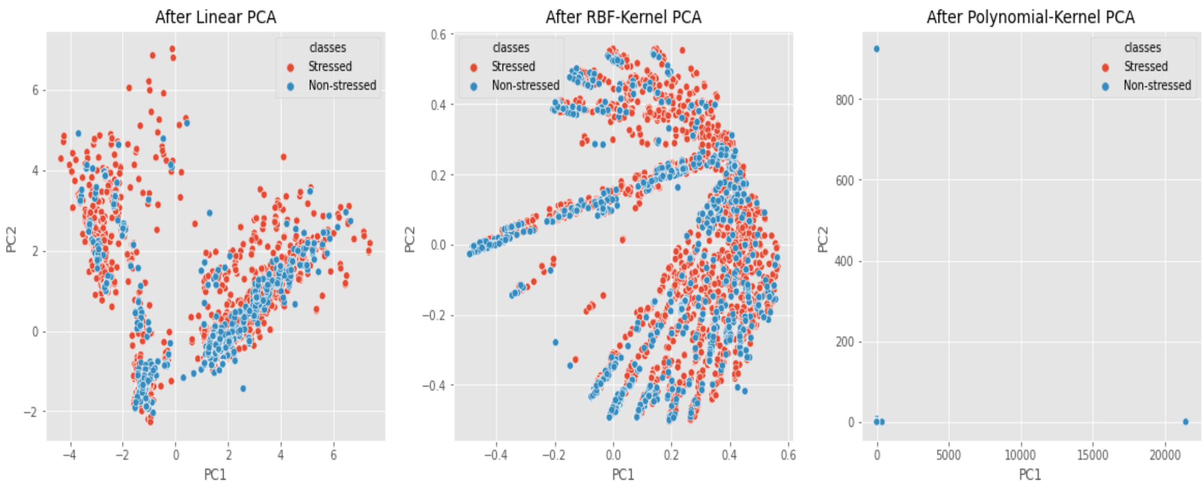


Figure 3: The projection of Dataset D2 on two principal components after different PCA techniques

5.2. Evaluation of the Proposed STSD Model

For evaluating the proposed STSD model's performance, many experiments are implemented on four datasets such as D1, D2, D3, and D4. All the results are recorded following the 10-fold cross validation of the model. Furthermore, the statistical significance of the results is analysed by conducting *one-sample Wilcoxon signed rank test* (Demšar, 2006). The null and alternate hypothesis for the *one-sample Wilcoxon signed rank test* is given as:

H_0 : The mean performance of other models is equal to the mean performance of the proposed STSD model.

H_1 : The mean performance of the other models differs from the mean performance of the proposed STSD model.

The decision of rejecting the null hypothesis is determined when observations of a proposed model are analysed with other models. In this work, the rejection of the null hypothesis is done with a $p-$value ≤ 0.05, inferring that the recorded results are statistically significant (Demšar, 2006).

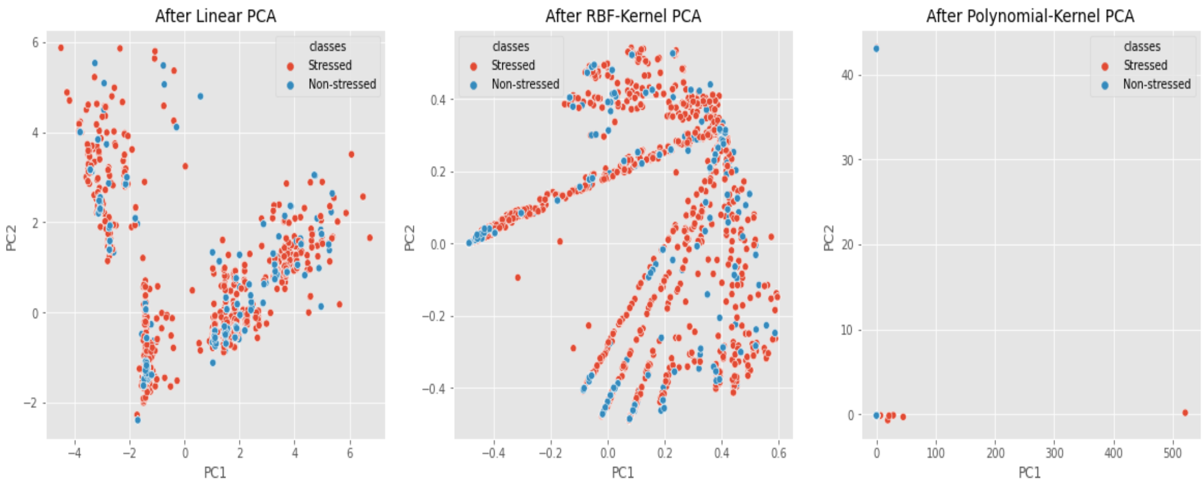


Figure 4: The projection of Dataset D3 on two principal components after different PCA techniques

Figure 5: The projection of Dataset D4 on two principal components after different PCA techniques

Table 3 stores the results of the experiments executed on the baseline models using the four datasets of D1, D2, D3, and D4 with original features. Initially, experiments are conducted on baseline models of Logistic Regression (LR), Support Vector Machines (SVM), Random Forest (RF), and Naïve Bayes (NB) along with the proposed STSD model. In Table 3, it is observed that the proposed STSD performs better than all other models on all the four datasets. The STSD model delivers an accuracy of 77.63% on dataset D1 and records an F1-score of 0.762. The SVM performs second best with respect to accuracy but records a low F1-score when compared with the proposed model. In the experiments on dataset D2, the proposed STSD model has leading performance with 76.79% accuracy and F1-score of 0.770. Here, SVM has the second best performance with 73.94% accuracy and an F1-score of 0.761. In the experiments on datasets pertaining to the third wave, D3 and D4, the STSD outperforms other baseline models. The STSD achieves an accuracy of 77.24% and an F1-score of 0.792 in experiments on dataset D3. While SVM and LR are ranked second and third in terms of both accuracy and F1-score. Also, on dataset D4, STSD records an accuracy of 76.75% and an F1-Score of 0.780, outperforming all other baseline models. Hence, when implemented with original features, it is concluded that the proposed STSD model delivers better performance when compared with all other baseline ML models.

Table 3
Performance results for experiments on all four datasets - with original features

Datasets	D1		D2		D3		D4	
Models/Measures	Accuracy (%)	F1-Score	Accuracy (%)	F1-Score	Accuracy (%)	F1-Score	Accuracy (%)	F1-Score
LR	74.62	0.734	73.92	0.750	74.91	0.772	74.4	0.761
SVM	77.30	0.744	73.94	0.761	75.96	0.772	75.06	0.762
RF	74.40	0.734	72.01	0.755	73.80	0.780	74.30	0.761
NB	74.30	0.734	72.70	0.750	74.01	0.772	74.70	0.761
STSD	**77.63**	**0.762**	**76.79**	**0.770**	**77.24**	**0.792**	**76.75**	**0.780**

Table 4
Performance results for experiments on all four datasets - By applying polynomial kernel PCA

Datasets	D1		D2		D3		D4	
Models/Measures	Accuracy (%)	F1-Score	Accuracy (%)	F1-Score	Accuracy (%)	F1-Score	Accuracy (%)	F1-Score
LR	82.71	0.805	77.16	0.781	78.08	0.798	76.34	0.770
SVM	83.74	0.821	77.17	0.788	77.64	0.816	76.86	0.770
RF	82.2	0.880	76.8	0.799	78.1	0.807	76.4	0.785
NB	82.6	0.860	76.3	0.765	79.2	0.790	76.2	0.766
STSD	**86.82**	**0.926**	**83.01**	**0.855**	**86.3**	**0.907**	**82**	**0.889**

Table 4 records the performances of all the models in this work after the application of polynomial kernel PCA. From table 4, it is noted that the performance of all the models on all datasets improves significantly by applying polynomial kernel PCA. The proposed STSD model delivers a leading performance with accuracy of 86.82% and F1-Score of 0.926 on dataset D1, while SVM has second highest accuracy of 83.74% and RF has second highest F1-Score of 0.880. For dataset D2, the proposed STSD records a high accuracy 83.01% with a high F1-Score of 0.855, while SVM records the second highest accuracy of 77.17% and RF records the second best F1-score of 0.799. After implementation on the datasets concerned with the third wave, D3 and D4, the proposed STSD model exhibits better performance than any other baseline models. The STSD exhibits a high accuracy of 86.3% and an F1-Score of 0.907 on dataset D3, while NB records the second best accuracy of 79.2% and SVM gives the second best F1-Score of 0.816. In the case of dataset D4, the STSD leads with an accuracy of 82% and an F1-Score of 0.889, while SVM has the second best accuracy of 76.86% and RF recording the second best F1-Score of 0.785. Accordingly, it is concluded that the proposed STSD leads in accuracy and F1-Score when compared to all other baseline ML models considered.

The bar-plots representing accuracy for each of the classifiers when implemented with original features and with polynomial kernel PCA, are presented in figure 6. It is observed that the STSD has the highest accuracy in both cases-first, with original features, shown with red coloured bars; and second, with polynomial kernel PCA, shown with cyan coloured bars. And the STSD model exhibits better accuracy than any other baseline model in both cases for all four datasets. Similarly, the bar-plots for F1-Score for each of the models considered with original features and with polynomial kernel PCA are depicted in figure 7. And the STSD model exhibits a better F1-Score than all other baseline models in both cases-first, with original features, shown with red coloured bars; and second, with polynomial kernel PCA, shown with cyan coloured bars. This is true for all the four datasets.

5.2.1. Analyzing the effect of using dimensionality reduction with STSD

The low performance of the models in their original features is noted in Table 3. The dimensionality reduction technique of polynomial kernel PCA is applied because it groups classes of tweets, as seen from the results of preprocessing in section 5.1. This helps in better classification. The accuracy and F1-score results on experiments conducted in this work on all four datasets with original features are presented in table 3. While the results of experiments after applying polynomial kernel PCA are presented in table 4. The bar-plot in figure 6 depicts the comparison of accuracies exhibited by all the models when implemented with original features and polynomial kernel PCA, for all the four datasets. The bar-plot in figure 7 compares the F1-Scores obtained by all models when implemented with original features and polynomial kernel PCA. The cyan bars in the figure 6 denote the accuracy of

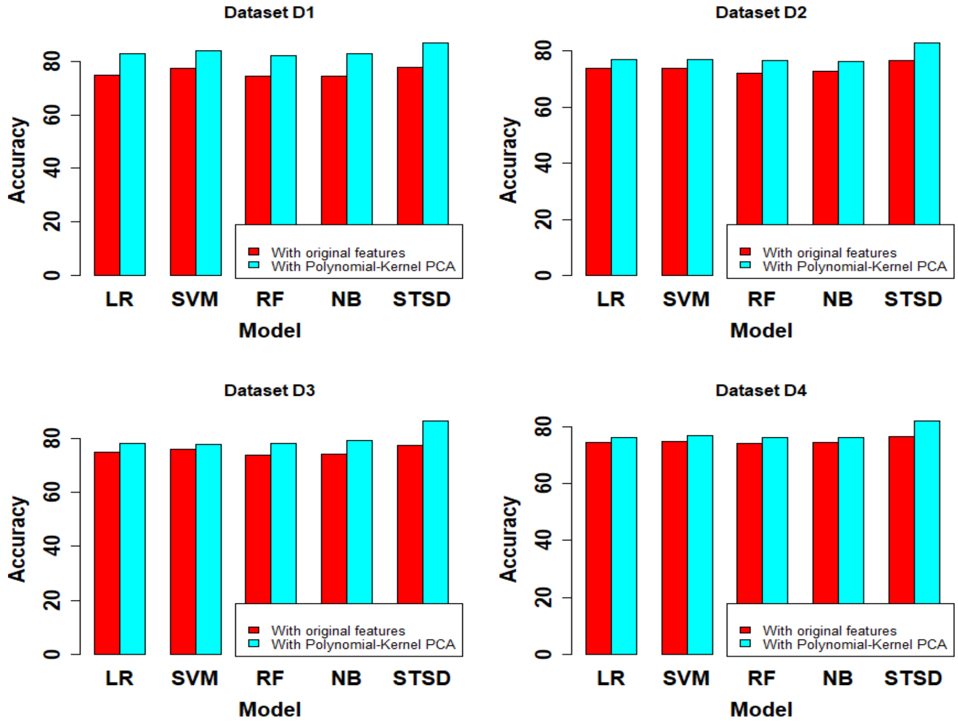


Figure 6: The bar-plots representing the accuracies of all models on the datasets D1, D2, D3, and D4

the models after the application of kernel PCA. Hence, it is observed that for all the models considered, the accuracy improves with the dimensionality reduction technique of polynomial kernel PCA. And the largest improvement is seen in the proposed STSD on all the four datasets. STSD with kernel PCA records at least 9.19% improvement in accuracy when compared to all the models implemented without PCA on dataset D1. While the same STSD records an improvement in accuracy by 6.22%, 9.06% and 5.25% when compared to other models implemented on datasets D2, D3, and D4, respectively. The least improvement in accuracy is observed when models are implemented on dataset D4. This is due to a lack of clear separation of classes after the application of polynomial kernel PCA, as observed from figure 5.

Similarly, the cyan bars in the figure 7 denote the F1-scores of the models after the application of the polynomial kernel PCA. It is observed that the F1-Score of all the models has seen an improvement after using the dimensionality reduction technique of polynomial kernel PCA. Moreover, the largest increment in F1-Score over all the datasets is observed for the proposed STSD model. When compared to all models implemented without PCA on dataset D1, STSD with kernel PCA improves F1-Score by at least 0.164 points. Whereas, the same STSD records an improvement of 0.085 points, 0.115 points, and 0.109 points when compared to other models implemented on datasets D2, D3, and D4, respectively. The least improvement in F1-Score is observed when models are implemented on dataset D4. This is due to a lack of clear separation of classes after the application of polynomial kernel PCA, as observed from figure 5.

5.3. Discussion

The proposed STSD model outperforms the baseline models on the datasets collected from two different time periods of COVID-19, as evidenced by the results presented in section 5.2.

The improvement in the performance of the proposed STSD approach with polynomial kernel PCA when compared to the other baseline models implemented with original features, is depicted in figures 8a and 8b. From figure 8a, it is noted that STSD with polynomial kernel PCA has a significantly better increment in accuracy compared to all models on datasets D1, D2 and D3 and a relatively small increment in accuracy on dataset D4. Figure 8b shows that in the case of dataset D1, a large improvement in F1-scores of the STSD model is seen when compared to all the baseline

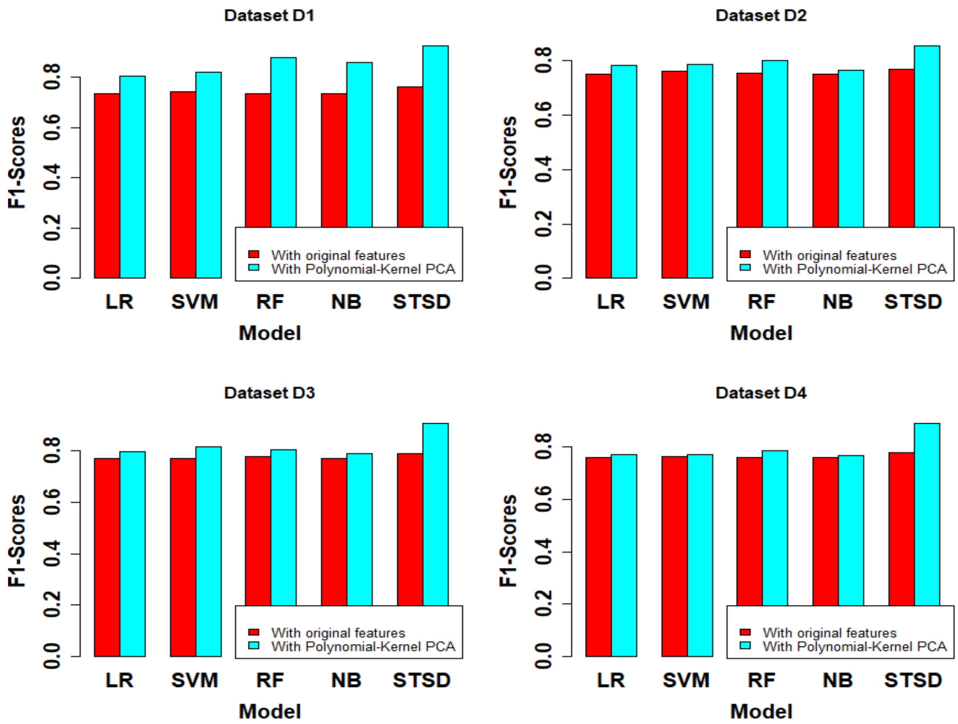


Figure 7: The bar-plots representing the F1-measure values of all models on the datasets D1, D2, D3, and D4

models. Whereas, in the case of dataset D2, the improvement in performance of the proposed STSD with polynomial kernel PCA is relatively smaller than the improvement recorded with all other datasets.

It is concluded that with the use of sarcasm information, STSD outperforms other baseline models. Furthermore, with the effective use of the PCA techniques, a better separation of classes is achieved. This resulted in good improvements in performances by all the models. And the proposed STSD model recorded the highest magnitude of enhancement in performance. Hence, STSD with polynomial kernel PCA is the better model for tweet-level stress detection, by enhancing the utilisation of tweet-content.

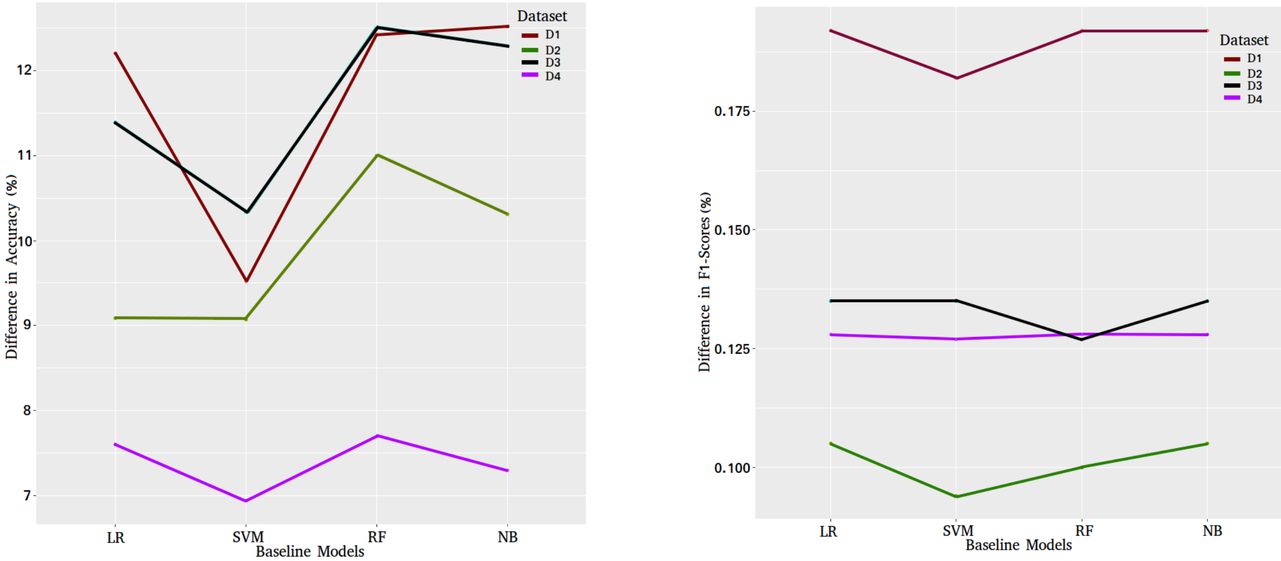
Finally, the limitations of the proposed STSD model are presented below:

- The proposed model will work only if the tweets contain textual content information.
- As text is the major source for extracting the features, the proposed model would perform well only for tweets that contain sarcasm or its forms present in textual content rather than embedded in media other than text.

6. Conclusion and Future Work

Psychological stress has emerged as a major global health problem, and social media-based stress detection has caught the attention of numerous researchers. In this work, for detecting tweet-level stress, a novel sarcasm-based tweet-level stress detection (STSD) classifier has been developed for availing the information related to sarcasm present in the tweet-content.

The principle of the STSD model is to minimise the loss for non-sarcastic tweets while maximising the loss for sarcastic tweets. Subsequently, a theorem is framed to prove the loss function of the proposed STSD model as the special form of the standard logistic loss. The experiments are conducted on four different datasets, with datasets D1 & D2 collected during the second wave of the COVID-19 pandemic and the datasets D3 & D4 collected during the third wave of the pandemic in India. A thorough preprocessing is performed and an appropriate dimensionality reduction technique of polynomial kernel PCA is applied to all the datasets so as to enhance the performance of the



(a) Accuracy (b) F1-Scores

Figure 8: Performance differences noted between baseline models with original features and STSD model with polynomial kernel PCA

models. From the experimental results, it is noted that the proposed STSD outperforms all the other baseline models in both the cases of the implementation with original features and the implementation after polynomial kernel PCA. The STSD model with polynomial kernel PCA records accuracies of 86.82%, 83.01%, 86.3%, and 82%, with datasets D1, D2, D3, and D4 respectively. Also, STSD with polynomial kernel PCA achieves an improvement in accuracy of at least 9.19%, 6.22%, 9.06%, and 5.25% when compared to the baseline models implemented without PCA on datasets D1, D2, D3, and D4 respectively. Moreover, STSD with polynomial kernel PCA records F1-Scores of 0.926, 0.855, 0.907, and 0.889 with datasets D1, D2, D3, and D4 respectively. Also, STSD with polynomial kernel PCA achieves an improvement in F1-score by at least 0.164 points, 0.085 points, 0.115 points, and 0.109 points when compared to all other baseline models implemented without PCA on datasets D1, D2, D3, and D4 respectively. In addition, all the models considered in this work record improved performance when implemented with polynomial kernel PCA as compared to their implementation without PCA. Wherein, the proposed STSD shows the highest improvement in accuracy and F1-score when compared to the increment recorded by all the other baseline models.

The future direction for this work is to develop multi-task learning models for utilising the clues from related classification tasks in detecting tweet-level stress. The other direction is employing efficient deep learning architectures to solve the problem of tweet-level stress detection.

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

References

Agarwal, A., 2020. Ripple effect of a pandemic: Analysis of the psychological stress landscape during covid19. PsyArXiv .

Asrani, P., Eapen, M.S., Hassan, M.I., Sohal, S.S., 2021. Implications of the second wave of covid-19 in india. The Lancet Respiratory Medicine 9, e93–e94.

Baker, S.R., Bloom, N., Davis, S.J., Terry, S.J., 2020. Covid-induced economic uncertainty. Technical Report. National Bureau of Economic Research.

Boser, B.E., Guyon, I.M., Vapnik, V.N., 1992. A training algorithm for optimal margin classifiers, in: Proceedings of the fifth annual workshop on Computational learning theory, pp. 144–152.

Breiman, L., 2001. Random forests. Machine learning 45, 5–32.

Camp, E., 2012. Sarcasm, pretense, and the semantics/pragmatics distinction. Noûs 46, 587–634.

Chhatwani, M., Mishra, S.K., Rai, H., 2022. Active and passive social media usage and depression among the elderly during covid-19: does race matter? Behaviour & Information Technology , 1–12.

Chicco, D., Jurman, G., 2020. The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation. BMC genomics 21, 6.

Coppersmith, G., Harman, C., Dredze, M., 2014. Measuring post traumatic stress disorder in twitter, in: Eighth international AAAI conference on weblogs and social media.

De Choudhury, M., Counts, S., Horvitz, E., 2013a. Predicting postpartum changes in emotion and behavior via social media, in: Proceedings of the SIGCHI conference on human factors in computing systems, pp. 3267–3276.

De Choudhury, M., Gamon, M., Counts, S., Horvitz, E., 2013b. Predicting depression via social media, in: Seventh international AAAI conference on weblogs and social media.

Demšar, J., 2006. Statistical comparisons of classifiers over multiple data sets. The Journal of Machine learning research 7, 1–30.

Fast, E., Chen, B., Bernstein, M.S., 2016. Empath: Understanding topic signals in large-scale text, in: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 4647–4657.

Frenda, S., Cignarella, A.T., Basile, V., Bosco, C., Patti, V., Rosso, P., 2022. The unbearable hurtfulness of sarcasm. Expert Systems with Applications , 116398.

Gandhi, U.D., Kumar, P.M., Babu, G.C., Karthick, G., 2021. Sentiment analysis on twitter data by using convolutional neural network (cnn) and long short term memory (lstm). Wireless Personal Communications , 1–10.

Glaser, R., Kiecolt-Glaser, J.K., 2005. Stress-induced immune dysfunction: implications for health. Nature Reviews Immunology 5, 243–251.

Guntuku, S.C., Sherman, G., Stokes, D.C., Agarwal, A.K., Seltzer, E., Merchant, R.M., Ungar, L.H., 2020. Tracking mental health and symptom mentions on twitter during covid-19. Journal of general internal medicine 35, 2798–2800.

Han, J., Pei, J., Kamber, M., 2011. Data mining: concepts and techniques. Elsevier.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The elements of statistical learning: data mining, inference, and prediction. Springer Science & Business Media.

Huo, J., Turner, K., 2019. Social media in health communication, in: Social web and health research. Springer, pp. 53–82.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An introduction to statistical learning. volume 112. Springer.

Joshi, A., Bhattacharyya, P., Carman, M.J., 2017. Automatic sarcasm detection: A survey. ACM Computing Surveys (CSUR) 50, 1–22.

Kamvar, S.D., Harris, J., 2011. We feel fine and searching the emotional web, in: Proceedings of the fourth ACM international conference on Web search and data mining, pp. 117–126.

Kim, A., Song, Y., Kim, M., Lee, K., Cheon, J.H., 2018. Logistic regression model training based on the approximate homomorphic encryption. BMC medical genomics 11, 83.

Koller, D., Friedman, N., Džeroski, S., Sutton, C., McCallum, A., Pfeffer, A., Abbeel, P., Wong, M.F., Heckerman, D., Meek, C., et al., 2007. Introduction to statistical relational learning. MIT press.

KVTKN, P., Ramakrishnudu, T., 2021. A novel method for detecting psychological stress at tweet level using neighborhood tweets. Journal of King Saud University-Computer and Information Sciences .

Le, X.T.T., Dang, A.K., Toweh, J., Nguyen, Q.N., Le, H.T., Do, T.T.T., Phan, H.B.T., Nguyen, T.T., Pham, Q.T., Ta, N.K.T., et al., 2020. Evaluating the psychological impacts related to covid-19 of vietnamese people under the first nationwide partial lockdown in vietnam. Frontiers in psychiatry , 824.

Lee, J.A., Verleysen, M., 2007. Nonlinear dimensionality reduction. volume 1. Springer.

Lin, H., Jia, J., Guo, Q., Xue, Y., Huang, J., Cai, L., Feng, L., 2014a. Psychological stress detection from cross-media microblog data using deep sparse neural network, in: 2014 IEEE International Conference on Multimedia and Expo (ICME), IEEE. pp. 1–6.

Lin, H., Jia, J., Guo, Q., Xue, Y., Li, Q., Huang, J., Cai, L., Feng, L., 2014b. User-level psychological stress detection from social media using deep neural network, in: Proceedings of the 22nd ACM international conference on Multimedia, pp. 507–516.

Lin, H., Jia, J., Nie, L., Shen, G., Chua, T.S., 2016. What does social media say about your stress?., in: IJCAI, pp. 3775–3781.

Lin, H., Jia, J., Qiu, J., Zhang, Y., Shen, G., Xie, L., Tang, J., Feng, L., Chua, T.S., 2017. Detecting stress based on social interactions in social networks. IEEE Transactions on Knowledge and Data Engineering 29, 1820–1833.

Liu, P., Chen, W., Ou, G., Wang, T., Yang, D., Lei, K., 2014. Sarcasm detection in social media based on imbalanced classification, in: International Conference on Web-Age Information Management, Springer. pp. 459–471.

Martins, R., Almeida, J.J., Henriques, P.R., Novais, P., 2021. Identifying depression clues using emotions and ai., in: ICAART (2), pp. 1137–1143.

Marzouki, Y., Aldossari, F.S., Veltri, G.A., 2021. Understanding the buffering effect of social media use on anxiety during the covid-19 pandemic lockdown. Humanities and Social Sciences Communications 8, 1–10.

Mattioli, A.V., Sciomer, S., Maffei, S., Gallina, S., 2021. Lifestyle and stress management in women during covid-19 pandemic: impact on cardiovascular risk burden. American journal of lifestyle medicine 15, 356–359.

Mukherjee, S., Bala, P.K., 2017. Sarcasm detection in microblogs using naïve bayes and fuzzy clustering. Technology in Society 48, 19–27.

Mundotiya, R.K., Yadav, N., 2021. Forward context-aware clickbait tweet identification system. International Journal of Ambient Computing and Intelligence (IJACI) 12, 21–32.

Nicola, M., Alsafi, Z., Sohrabi, C., Kerwan, A., Al-Jabir, A., Iosifidis, C., Agha, M., Agha, R., 2020. The socio-economic implications of the coronavirus pandemic (covid-19): A review. International journal of surgery 78, 185–193.

Pratama, B.Y., Sarno, R., 2015. Personality classification based on twitter text using naive bayes, knn and svm, in: 2015 International Conference on Data and Software Engineering (ICoDSE), IEEE. pp. 170–174.

Rajadesingan, A., Zafarani, R., Liu, H., 2015. Sarcasm detection on twitter: A behavioral modeling approach, in: Proceedings of the eighth ACM international conference on web search and data mining, pp. 97–106.

Satour, N., Benyacoub, B., El Mahrad, B., Kacimi, I., 2021. Kpca over pca to assess urban resilience to floods, in: E3S Web of Conferences, EDP Sciences. p. 03005.

Sundararajan, K., Palanisamy, A., 2020. Multi-rule based ensemble feature selection model for sarcasm type detection in twitter. Computational intelligence and neuroscience 2020.

Sv, P., Lathabhavan, R., Ittamalla, R., 2021. What concerns indian general public on second wave of covid-19? a report on social media opinions. Diabetes & metabolic syndrome 15, 829.

Thelwall, M., 2017. Tensistrength: Stress and relaxation magnitude detection for social media texts. Information Processing & Management 53, 106–121.

Wang, Y., Wang, Z., Li, C., Zhang, Y., Wang, H., 2020. A multitask deep learning approach for user depression detection on sina weibo. arXiv preprint arXiv:2008.11708 .

Xiong, J., Lipsitz, O., Nasri, F., Lui, L.M., Gill, H., Phan, L., Chen-Li, D., Iacobucci, M., Ho, R., Majeed, A., et al., 2020. Impact of covid-19 pandemic on mental health in the general population: A systematic review. Journal of affective disorders 277, 55–64.

Xu, J., Wang, F., Ai, J., 2020. Defect prediction with semantics and context features of codes based on graph representation learning. IEEE Transactions on Reliability .

Xue, Y., Li, Q., Jin, L., Feng, L., Clifton, D.A., Clifford, G.D., 2014. Detecting adolescent psychological pressures from micro-blog, in: International Conference on Health Information Science, Springer. pp. 83–94.

Xue, Y., Li, Q., Zhao, L., Jia, J., Feng, L., Yu, F., Clifton, D.A., 2016. Analysis of teens' chronic stress on micro-blog, in: International Conference on Web Information Systems Engineering, Springer. pp. 121–136.

Zhang, Y., Lyu, H., Liu, Y., Zhang, X., Wang, Y., Luo, J., et al., 2021. Monitoring depression trends on twitter during the covid-19 pandemic: Observational study. JMIR infodemiology 1, e26769.

Zhao, L., Jia, J., Feng, L., 2015. Teenagers' stress detection based on time-sensitive micro-blog comment/response actions, in: IFIP International Conference on Artificial Intelligence in Theory and Practice, Springer. pp. 26–36.