GAUGING STRESS, ANXIETY, DEPRESSION IN STUDENT DURING COVID-19 PANDEMIC

ASTHA SINGH ∗ AND DIVYA KUMAR †

Abstract. During the beginning of COVID-19 pandemic, studies came across the world concerning with health issues. Researches began to find the repercussions of the virus. The virus was found to be versatile as it changes its nature and targets the lungs of a person. Later, it was seen an astonishing massacre around the world due to the virus. Many people have lost their life but many more people are still suffering with bad psychological state. Researchers began to research on the nature virus but very few researches were made on the other side-effects of this pandemic. One such crucial subject to attend in contemporary world is the effect of COVID-19 on psychological state in general population. This side-effect may lead to raise an alarming situation in future that could result in more death cases. The proposed paper presents a study on the detection of stress and depression in people caused by the pandemic. The proposed methodology is based on perceived questionnaire method through which people’s responses are recorded in the form of text. COVID victims have been interrogated against a set of questions and their responses are recorded. The methodology performs text mining of their responses that also include the people’s reaction from social networking sites. The text processing of people’s responses is done by natural language processing (NLP). NLP is used to interpret textual facts into meaningful segments that must be understandable to machine. The refined data has been transformed into PSS (perceived stress scale) scaling factor that ranges from 0 to 4 showing various level of stress. The proposed system utilized artificial intelligence in which naïve Bayes classifier, K-nearest neighbor (KNN), Decision tree and Random forest algorithms are applied to predict the emotional state of a person. The proposed system also uses data from social networking site for testing purpose. The model successfully shows a comparative study of such three classifiers for the classification of stress level into stress, anxiety and depression.

Key words: Text processing, perceived questionnaire method, Natural language processing, naïve Bayes, K-nearest neighbor, principle component analysis

AMS subject classifications. 68T05

1. Introduction. Stress, depression, and anxiety (SAD) are the most common psychological adversity that affects human life indefinitely. The bad mental state may excite many other diseases in the human body. In India, the reasons for stress or depression are mostly related to personal or professional life. But the current situation in which COVID-19’s second wave overwhelmed India’s healthcare system resulted in a sudden rise of infected people, which goes about 4 lakh per day [1]. The number of deaths reported was around 4000 per day after March 2021. The rate of spread has seen confounding effects due to the heavier population density in India. India till now reported about 331,909 death cased since Jan 2020. With the huge death reports, India became the second country that suffered from the most infected cases and deaths. With the huge number of death reports due to COVID-19, India is also suffering from a psychological imbalance that may exist for a long duration. According to WHO [2] (world health organization), psychological adversity has also been affected by quarantine for several months, during which people became less interactive with their society. COVID-19 also affects India’s economy badly, due to which financial crises have been seen. Common people in India suffered financial losses due to unemployment also. All these unpleasant events contribute to mental disruption in Indian society. The WHO shared their concern on the consequences of obstruction in daily routines and livelihoods of people in Indian society. Several deaths in the second wave [3] caused rapid inclination in loneliness, anxiety, depression, insomnia, suicidal cases, etc. The lockdown also ignited domestic violence, as revealed by one of the studies by Abramson et al. According to Loiwal et al. [4], the second wave of the coronavirus has caused a 20% increase in mental disease, according to the Indian Psychiatric Society. According to a report by Jayakumar et
al. in Kerala, India, banning alcohol sales during the lockdown has also revealed mental disease among alcohol addicts. The strength of the psychological imbalance among persons after the second wave has not yet been carefully examined because of the lockdown and quarantine regulations. On social networking platforms where users share their experiences, stress and sadness are easily tracked. The study of stress detection using facial images or video datasets may not seem feasible for the people affected by the pandemic. Such methods require a lab setup and volunteers directly or indirectly affected by COVID. The people spend most of their time at home due to protocols of lockdown and quarantine. They mostly shared their experiences on social media. Suicidal cases have also seen an increase in India. People show reluctance in making gatherings or attending any experimental subject. The experimental setup based on facial images or video datasets is complex and expensive.

The proposed method applied the perceived questionnaire method to determine the psychological state level. This method is based on the interrogation of COVID victims. The method records people's responses in the form of text data. Other textural data from a user who may not appear in an interrogation session has been collected from social networking sites. The responses are compiled and analyzed using the natural processing language (NLP). The NLP breaks down each patient response and correlates them in perceived scaling factor ranges from 0 to 4. The perceived stress scale records people's perceptions and converts them in to score points. The COVID victims are questioned following standard protocol, and their answers are recorded. To estimate the severity of the stress level, the proposed method used the naive Bayes classifier, K-nearest neighbor (KNN), Decision tree, and Random forest algorithms. As a result, the technique divides the reactions into stress, anxiety, and sadness. The classification results for stress, anxiety, and depression have been calculated in a comparative classification study that has been evaluated. The pre-processing techniques are utilized and consider best as it generates effective filtered response relevant for the detection of stress, anxiety and depression.

The following sections comprise the remaining portion of the paper:- Natural language processing, the perceived stress scale, and naive Bayes algorithms are discussed in Section 2 of the proposed methodology. The results section in Section 3 illustrates how well the algorithm worked. We'll talk about the work's conclusion in Section 4. References are included in the final section.

2. Proposed methodology. In the proposed methodology, first the dataset of questionnaire has been taken from standard Microdata Library[9]. This dataset specifically prepared for COVID victims in six different states of India including Jharkhand, Rajasthan, Uttar Pradesh, Andhra Pradesh, Bihar, and Madhya Pradesh. It covers indications related to agriculture, migration, rural, labour markets, consumption patterns, access to relief and healthcare etc. The datasets has been recorded in the may-2020 after the outbreak of second wave of COVID-19 in India [16]. The dataset is sponsored by worldbank [17]. The dataset contain questionnaire for stress detection targets questionnaire related to daily-life activities of COVID victims. The responses from the peoples are collected through personal interrogation, online survey and social media where questionnaire set has been posted. The proposed method also takes textural responses from general post from social networking sites. The correlations of such public posts have been checked with the questionnaire dataset. These responses must correlates with the questionnaire set otherwise the responses are discarded. The responses are based on perception of peoples. Then, the proposed model applies natural processing language [14] (NLP) that performs pre-processing and analysis of textural responses. Then the refined data will be mapped with the PSS scale that ranges from 0 to 4 defining various levels of stress. Then naïve Bayes classifier is used that perform classification of PSS responses based probabilistic approach. The workflow of the proposed system has been described in figure 2.1.

As depicted in figure 2.1, the responses are captured from several peoples over a set of 500 questionnaire data. Each person replied based on his/her perception.

2.1. Natural language processing (NLP). Natural language processing has been applied to the various textural responses recorded from the public domain. The NLP applies syntax analysis and semantic analysis of the data. In syntax analysis, operations like lemmatization, morphological segmentation, word segmentation, sentence breaking, parsing, and stemming are involved. Lemmatization is converting a complex sentence or word into a simple form. Morphological segmentation is used to segregate the sentence into distinct units. Parsing performs grammar checks for the textural responses. Sentence breaking is the breaking of a large sentence into words. Stemming involves the conversion of words into their root form. The semantic analysis
NLP performs a meaning check on the textural data. It analyzes the sense of the words in the responses. NLP applies from social networking sites to obtain textural data and correlates it with the existing questionnaire. NLP eradicates all the ambiguity, extra spaces, duplicity, etc., from the dataset and prepares it to correlate with the PSS scaling factor. Figure 2.2 shows the basic working of NLP. The NLP process is used to interpret the meaning of each word and letter of the response taken against questionnaire dataset.

Figure 2.2 shows the character filtration of responses in which several modules are used. The processed language is divided into training and testing set for the validation. The modules used in the character filtration are described in figure 2.3.

3. NLP gives a refined output that is further mapped with PSS scaling factor.

3.1. PSS (Perceived stress scale). An effective psychological tool for assessing stress perception is the Perceived Stress Scale (PSS) [19]. It is a gauge of how stressful certain circumstances are in a person’s life. Things are made to change how the unruly, unpredictable, and overburdened respondents feel about their lives. Numerous direct questions about the current state of stress are also included on the scale. The PSS is intended to be applied to community samples that have completed at least a high school degree. Things are simple to comprehend, and some responses are also simple to comprehend. The questions are also generic, thus no substance is directed at any inferior group. PSS questions inquire about emotions and thoughts from the previous month. Respondents were questioned about how frequently they experienced each feeling. The PSS
might be between 0 and 40. Low stress is estimated to be between 0 to 30. Scores between 14 and 26 are regarded as moderate stress, whereas scores between 27 and 40 are regarded as high stress. The scores are converted into 5 PSS scaling factors by the suggested methodology. The PSS response scores are calculated using the terms 0, 1, 2, 3, and 4. Each score value identifies a particular stress level in an individual.

0 : - Stress never occurs.
1 : - Almost never, yet there is still a risk for tension.
2 : - Stress occurs sometimes and forms anxiety.
3 : - Stress occurs fairly often considered into moderate anxiety
4 : - Stress occurs very often is considered into depression.

PSS has 14 total scores, which includes the 5 answer points. The total PSS score is created by reversing the PSS scores 4,5,6,7,9,10, and 13 such that 0=4, 1=3, 2=2, 3=1, and 4=0, then adding together all 14 values. A high PSS score indicates a high level of stress.

**Evidence for Validity.** Higher PSS scores cause the following association:

1. failure to quit addiction
2. failure in curing diabetics
3. Health issues, depressive symptoms
4. Continuous headache and body pain.

### 3.2. Classification algorithms.

A comparative result of these three classifiers is calculated and shown in the result section.

#### 3.3. Naïve Bayes.

The Bayes theorem has been used in the nave Bayes classifier to determine the likelihood that a feature vector belongs to a category that includes sadness, stress, and anxiety. The classifier is used to analyze the data it receives from the PSS scaling factors. The classifier is used to map the chances that various feature spaces belong to a class. The feature that has the highest probability is categorized into a particular class. The dataset has been split into training and testing portions that are each 70:30.

The equation of Bayes theorem is defined as:

\[
P(Y|K) = P(K|Y)P(Y)/P(K)
\]

Here,

1. \(P(Y|K)\) is the probability of \(Y\) such that event \(K\) is already true.
2. \( P(Y) \) prior probability of classes.
3. \( Y \) and \( K \) are the two events.
4. Probability of predictor given class

The naive Bayes classifier compares the probabilities of belongingness of various feature space in a class. Feature that has maximum probability is classified into a specific class. The dataset has been divided into 70:30 for training and testing respectively.

3.4. **K-nearest neighbor (KNN)**. The KNN algorithm is based on a distance metric assessed by the degree of linguistic similarity. In PSS records, the distance algorithm has been used, and the textural data has been divided into categories for stress, depression, and anxiety.

**Step 1:** Input PSS records

**Step 2:** The similarity between two selected textural data has been measured by using the following equations:

\[
I(A) = \sum_{m=0}^{m} \sum_{n=0}^{n} M_m M'_y
\]

\[
D_1(I_x, I_y) = \sum_{m=1}^{N} |M_m - M'_y|
\]
Astha Singh, Divya Kumar

3.5. Decision Tree. Decision tree algorithm takes the input PSS record and transforms it into a tree having root node, parent node and the child node. Each node shows some information. Information of parent node may depend on the information of child nodes. The classification of entire information is represented by leaf node. The decision of stress, anxiety and depression is hierarchically decided by information [20] of non-leaf nodes.

The emotion recognition model uses a dataset of around 37K images with each image being of the size 48x48 as we can see in the figure 2.2, these images are in the black and white mode, the name of the dataset is FER dataset. The model made takes the ELU activation function to process the input data along with the batch normalization technique to make the model run for lower amount of time in each epoch, absence of batch normalization can make each epoch of the model to run for at least 90 minutes each. The model is later is flattened with the use of same activation function and then later the output layer in the model has been taken as the SoftMax function which helps us in making a fully connecting layer. The model uses the Maxpooling function to pool the best feature onto the next layer as seen in the figure 2.3.

3.6. Random Forest. Random forest is the combination of multiple decision trees. It is used to represent data into various decision trees and combined their information to produce single output. The leaf node of each tree shows the target decision i.e. stress, depression and anxiety. The decision of all the trees is combines to
obtain an equivalent classification of data. Each single tree in the model separately predicts for the classification of data. These predictions are combined together to obtain the final output of the model. The decision trees operate as an ensemble.

4. Result discussion. The results are carried out on the standard dataset taken from the Microdata library [9] in which a questionnaire has been prepared and asked from 6 different states of India. The questionnaire was taken in the form of text data. Then, NLP is applied to conduct pre-processing. The pre-processed data is further converted into PSS scaling factors. There are other noise removal techniques such as filtration methods depending on datasets. The utilized pre-processing techniques in the proposed model is effective as the questionnaire dataset is concerned. The questionnaire responses contained in terms of textual data that needs to be normalized. It has been done by the proposed pre-processing techniques in low cost and high effectiveness shown by the robustness. Table 4.1 shows the sample of the questionnaire and the respective PSS records scaling from 0 to 4.

Table 4.1 shows the samples of questions taken from the standard dataset. The responses are transformed into PSS scaling factors. The median and mode has also been calculated for the PSS ratings. These PSS records are classified into stress, anxiety and depression. The classification task has been performed four types of classifier i.e. Naïve Bayes, K-nearest neighbour, Decision tree and the random forest.

Table 4.2 shows the confusion matrices obtained by all the four classifiers that is used to classify anxiety, depression and stress.

The below equations are used for the calculation of accuracy rate, precision rate, recall and F1 score:

\[
\text{Accuracy Rate} = \frac{\text{Diagonal sum (MP)}}{\text{Total number of particles}}
\]

\[
\text{Error Rate} = 1 - \text{Accuracy Rate}
\]

\[
\text{Precision Rate} = \frac{\text{MP}}{\text{MP + NF}}
\]

\[
\text{Specificity} = \frac{2 \times (\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})}
\]

where
- MP indicate - (True positive) = Matrix diagonal
- NF indicate - (False Negative) = Consistent row for class
- FM indicate - (False Positive) = Corresponding column for class
- NF indicate (False Negative) = Sum of all column and row
Fig. 4.1: Graph showing confusion matrices for SAD Naïve Bayes Classifier

Fig. 4.2: Graph showing confusion matrices for SAD KNN

Fig. 4.3: Graph showing the confusion matrices for SAD Decision tree
Table 4.1: Sample of questionnaire and the related PSS response (rated in a scale of 0-4) as reported from people

<table>
<thead>
<tr>
<th>Source of Stress</th>
<th>Never (0)</th>
<th>Almost Never (1)</th>
<th>Sometimes (2)</th>
<th>Fairly Often (3)</th>
<th>Very Often (4)</th>
<th>Media</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Have your personal plans been changed or affected by the COVID in the last 7 days?</td>
<td>28</td>
<td>24</td>
<td>35</td>
<td>22</td>
<td>5</td>
<td>11</td>
<td>5</td>
</tr>
<tr>
<td>Did you do any work for pay during COVID at a job or business in the last 1 month?</td>
<td>14</td>
<td>18</td>
<td>25</td>
<td>21</td>
<td>14</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Think about your life for past 30 days, how do you think, was it a really tough time?</td>
<td>20</td>
<td>17</td>
<td>21</td>
<td>29</td>
<td>27</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Have you currently covered by any of the types of medical insurance or any health plan?</td>
<td>23</td>
<td>14</td>
<td>29</td>
<td>38</td>
<td>12</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Have you had any friend or member who is very close to you die from COVID-19 or recovered from 2nd wave of COVID?</td>
<td>8</td>
<td>5</td>
<td>18</td>
<td>30</td>
<td>28</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>How did you communicate with your family, friends and relatives during the last 1 month? Did you use any phone, text, email, or Internet</td>
<td>22</td>
<td>20</td>
<td>24</td>
<td>15</td>
<td>6</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

Figures 4.1–4.4 are showing confusion matrices for anxiety, depression and stress. These confusion matrices are generated by all the four classifiers used in the proposed scheme.

Table 4.1 is showing the result on the basis of confusion matrices for SAD. These confusion matrices are represented by all the four classifiers used in the proposed model.

Table 4.2 depicted for the result scores including rate of accuracy, error value, rate of precision, rate of recall and F1 score. These scores are calculated for the respective anxiety, depression and stress represented as A, D and S. The various classifiers have been shown in this table based on which the results have been calculated.
Table 4.2: Measure of accuracy, error rate, precision, recall and F1 score for anxiety (A), depression (D) and stress (S)

<table>
<thead>
<tr>
<th>Name of Classifiers</th>
<th>Psychological status</th>
<th>Value of Accuracy</th>
<th>Value of Error Rate</th>
<th>Data of Precision</th>
<th>Data of Recall</th>
<th>F1 score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naïve Bayes</td>
<td>A</td>
<td>0.738</td>
<td>0.27</td>
<td>0.523</td>
<td>0.658</td>
<td>0.658</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>0.851</td>
<td>0.28</td>
<td>0.574</td>
<td>0.798</td>
<td>0.798</td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>0.708</td>
<td>0.29</td>
<td>0.596</td>
<td>0.798</td>
<td>0.745</td>
</tr>
<tr>
<td>KNN</td>
<td>A</td>
<td>0.745</td>
<td>0.28</td>
<td>0.795</td>
<td>0.715</td>
<td>0.748</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>0.796</td>
<td>0.25</td>
<td>0.853</td>
<td>0.896</td>
<td>0.821</td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>0.854</td>
<td>0.28</td>
<td>0.841</td>
<td>0.889</td>
<td>0.856</td>
</tr>
<tr>
<td>Decision tree</td>
<td>A</td>
<td>0.874</td>
<td>0.248</td>
<td>0.896</td>
<td>0.854</td>
<td>0.895</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>0.824</td>
<td>0.241</td>
<td>0.813</td>
<td>0.896</td>
<td>0.865</td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>0.887</td>
<td>0.267</td>
<td>0.796</td>
<td>0.874</td>
<td>0.879</td>
</tr>
<tr>
<td>Random forest</td>
<td>A</td>
<td>0.752</td>
<td>0.224</td>
<td>0.745</td>
<td>0.897</td>
<td>0.785</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>0.789</td>
<td>0.296</td>
<td>0.758</td>
<td>0.874</td>
<td>0.796</td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>0.796</td>
<td>0.274</td>
<td>0.746</td>
<td>0.752</td>
<td>0.798</td>
</tr>
</tbody>
</table>

Fig. 4.4: Confusion matrices for anxiety, depression and stress for Random Forest

Figure 4.5 shows the graphical comparison of accuracies obtained by all four classifiers for stress, depression, and anxiety. Figure 12 shows that the highest accuracy has been observed in the decision tree classifier compared to other studied classifiers. The relevance of the study is to prevent suicidal cases and improve the treatment policy of the patients having mental disorder. The questionnaire dataset is effective as it contains the exact patient’s responses and easy to interpret using tools and techniques. The limitation of the model is its static nature as the responses against the questionnaire dataset are static and it may not be effective as other dynamic dataset such as video-based interrogation, audio responses etc. The future of the model is in the medical science where it helps to reduce cost for the detection of mental disorder. Various other language filtration processing tools can be utilized in the place of NLP but the NLP method is found to be more effective as per the textual questionnaire responses is concerned of the proposed model. The frequency analysis is very crucial as it give details about intensity information based on statistical scores. The frequential information shows the exact
Fig. 4.5: Graphical representation of accuracy analysis for stress, depression and anxiety variation of psychological activities that helps to reach to the accurate detection of the stress, depression and depression.

5. Conclusion. The proposed work conducted an experiment to identify stress, depression, and anxiety using four different types of classifiers. The model successfully identifies how people’s psychological states relate to the COVID-19 pandemic. The dataset for the model is a collection of questionnaires that were created for the COVID second wave. The suggested system uses offline and online questioning to gather survey responses from various people. Utilizing natural processing language, the responses are analyzed and processed. PSS scaling factors have been created from the processed data. The data was then categorized into stress, depression, and anxiety using classification algorithms. The model is effective for the detection of stress, depression, and anxiety, according to the results section. The intensity level of the adverse psychological disorder is used to distinguish among stress, anxiety and depression. The proposed model concludes that the decision tree classifier is found to generate more accuracy as compared to other classifiers.

REFERENCES

Edited by: Vinoth Kumar  
Received: Jun 4, 2022  
Accepted: Dec 5, 2022