MENTAL HEALTH ANALYSIS USING NATURAL LANGUAGE PROCESSING

[1]K.DEEPA, [2]C.RANJEETH KUMAR, [3]M. KALEEL RAHMAN, [4]E. DERRICK GILCHRIST

[1]Professor, Department of Information Technology Sri Ramakrishna Engineering College, NGGO Colony Coimbatore, Tamil Nadu - 641022, India

.[2]Assistant Professor, Department of Information Technology Sri Ramakrishna Engineering College, NGGO Colony Coimbatore, Tamil Nadu - 641022, India

[3,4] Student, Department of Information Technology Sri Ramakrishna Engineering College, NGGO Colony Coimbatore, Tamil Nadu - 641022, India
E-mail: deepak9799979@gmail.com

ABSTRACT

Many people all over the world are depressed and are completely unaware of it. Depression is a mental illness in which a person is constantly unhappy and loses interest in almost everything. Depression can result in self-harm or even suicide. People can, thankfully, recover from depression with the help of therapy and medication. When a person's depression is detected early, his or her recovery will be greatly aided. Our project's main goal is to detect the depression in users' speech while also providing assistance for depression recovery. Nlp models such as word embedding and tone analyzer are used to detect depression, and recovery guidance given to the patient by providing the consultant details in their surroundings.

Keywords: *CNN, Word Embedding. Tonal Analysis, Depression Detection*.

1. INTRODUCTION

The capability of an individual to judge their own personal strengths and weaknesses and capable of identifying what is the maximum stress that they can overcome in their life is stated as Mental Health. If a person is mentally stable then he/she can do all the day to day work with full satisfaction and they we will be always productive. When someone's mental health is disrupted, they can't go about their daily lives in peace. They will experience intense, long-lasting sadness or despair. It can have a significant impact on a person's social behaviour and daily activities.

Depression is a common illness worldwide, with an estimated 3.8% of the population affected, including 5.0% among adults and 5.7% among adults older than 60 years . Approximately 280 million people in the world have depression.

People who are under depression are recognized by their behavior and also through their communication with others. Early recognition of depression is very crucial and greatly aids them to recover from their mental state. Without proper guidance depression may lead to self harm or even suicide. People who are under depression are recognized by their behavior and also by analyzing their communication with others. It is found that by analyzing their speech depression can be recognized more accurately than other methods. Depression can be recognised by speech recognition and tonal analysis.

The proposed work intends to implement a word embedding system for sentences used during speech, as well as a Convolutional Neural Network (CNN) model that uses spectrographic images of the tone for tonal analysis. The results of these two different models are ensembled to produce an effective output, which is discussed further in this paper.

2. LITERATURE REVIEW

Nowadays, social media is a popular platform for people to share their thoughts, feelings, and so on. This is the most effective method for obtaining information about a person

and determining whether or not they are depressed. A hybrid model that can detect depression by analyzing user textual posts has been implemented. A binary classification model that predicts depression or not is utilized for this purpose. Deep learning algorithms were trained using training data, and their performance was assessed using test data from the reddit dataset, Early Detection of Depression in CLEF eRisk 2019. Initially, feature extraction for the considered dataset is performed. TrainableEmbed Features, GloveEmbed Features, ord2VecEmbed Features, FastextEmbed Features, and Metadata Features are the five types of extracted features. These embedded features are fed to Bidirectional Long Short Term Memory (BiLSTM) layer and output is predicted [1]. Creating a mobile application and connecting it to the YouTube API allows us to efficiently track the search history on YouTube. The history is recorded and saved at the end of each search session. This data is subjected to sentimental analysis, and depression is predicted [2]. Given the complexities involved in the identification and treatment of mental disorders, an automated approach to identifying mental illnesses based on web mining and emotion analysis is used. Data from the Twitter social media platform is considered, and tokenization is performed with the NLTK Tweet tokenizer. The most frequently occurring words are extracted and fed into two different models. The first is CNN, which utilizes the CNNWithMAX, MultiChannelCNN, and MultiChannelPoolingCNN approaches. The other technique is RNN, which employs Bidirectional LSTM with attention and Context-aware Attention methodologies. It has been discovered that CNN-based models outperform RNN-based models. Models with optimized embeddings were able to maintain performance while also being generalizable. Models with optimized embeddings managed to maintain performance with the generalization ability [3]. The Reddit dataset is also used for depression identification among online users. Initially the lexicon of terms that are more common among depressed accounts is identified. Then stemming is applied in order to reduce the words to their root form and group similar words together. The strength and effectiveness of the combined features (LIWC+LDA+bigram) are most successfully demonstrated with the Multilayer Perceptron (MLP) classifier resulting in the top performance for depression detection [8]. In social media feeds not only textual data are considered but also images and selfies associated with those posts. The text data from the posts are separated and passed into a BERT model for word embedding whereas the visuals are passed into a CNN model for feature extraction. These features are combined into tensors for training the classification model. This approach had better performance compared to previously provided other approaches [5]. IIt is also possible to combine video, audio, and lexical features to classify depression. The only limitation is that this kind of data is not readily available, which led to proceeding with random interview videos which are separated by topics and fed to classification models. Root mean square error (RMSE), Mean absolute error (MAE), Pearson correlation coefficient (CC) and F1-score are the evaluation metrics that could be used for this approach [6]. Online activities like not only posts but also other features like user profile are also used for depression detection and it majorly analyzes the online behavior of depressed people. Social network features, user profile features, visual features, emotional features, topic-level features, and domain-specific features are the features extracted from online activity. The datasets are divided into two classes Depression Dataset D1 and Non - Depression Dataset D2. These Dataset are constructed using available twitter API's around the world which are also used for monitoring social media activity in twitter. Naive Baison and MSNL (Multiple Social Networking Learning) are the two types of classification models used [7]. Evaluating a depression detection model also plays an important role. There are 2 kinds of approach for evaluation : sequential and non sequential approaches. Initially, the features are extracted from the social media posts. The sentences or phrases from the posts are segmented and preprocessed by removing stop words and other delimiters. The frequency of unique words are calculated if the word is associated with depression or just common words are calculated and passed as features to the model.

In a non sequential approach, The features from the posts are extracted without any order and passed to an SVM for classification. But in a sequential approach the features are concatenated into a vector based on the time of the posts posted and passed into a recurrent neural network for classification. ERDE (Early Risk Detection Error and), Latency and Latency-weighted F1 is used for evaluation. Latency-weighted F1 outperformed others by displaying clear discrimination among models [4]. Audios can be represented as spectral representation and used as features for training deep learning models. Mel spectrograms are widely used for analyzing audio in many studies. The features from the audio are represented as Mel spectrograms and given as input for training in deep learning models like CNN. CNN architectures are used widely as Mel spectrograms are more like snapshots of Audio. Instead of using normal or any other spectrogram representation, Mel spectrograms are used as they can interpret more features from the audio data and also it is convenient to use Mel spectrograms to train CNN architectures like Resnet using transfer learning[9]. Humans express their emotions and state of mind by exposing them in their speech and facial expressions. Speech is considered to be the best way to detect a human's state of mind. Over the years by analyzing many studies and experiments conducted by researchers Mel spectrogram is considered as one of the optimal ways of representing the audio features. As they interpret the change in amplitude in human voice accurately. So they are widely used in emotion recognition [10].

3.ALGORITHM

This study is proceeding with two different approaches for detecting a person's depression state.
1.Semantic Analysis
2. Convolutional Neural Network based Tonal Analysis
The first approach is based on semantics. In this method, the depressed and non-depressed sentences are stored in a CSV file and fed into the Semantic model for training.[11-15] Approach 2 trains the CNN based tonal analysis model by converting the person's speech to a MEL Spectrogram image and passing it to the CNN model.

A. Semantic Analysis

Machine learning algorithms can work with 2-D arrays, where the rows and columns correspond to instances and features, respectively. In order to use these machine learning algorithms, we must convert the sentences into vector form, which is referred to as vectorizing the documents. In this we will be using a vectorizer and a classifier algorithm for detecting depression.

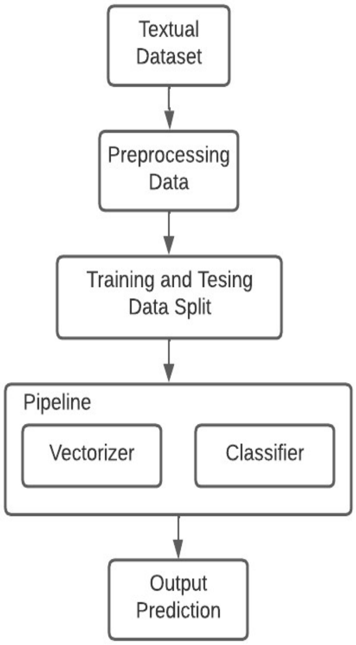


Fig. 1 Semantic Analysis

a) TF-IDF Vectorizer: The machine learning model cannot be trained directly on text data. As a result, converting text into a matrix or vector will allow the ML model to be trained. This is known as the feature extraction process. There are two types of vectorizers that are commonly used for text-to-vector conversion. The first is the Count vectorizer, and the second is the TF IDF vectorizer. The count vectorizer generates a matrix with the documents in the rows and the unique words from those documents in the columns. The cells contain binary values such as 0 and 1; if the word appears in the document, the value will be 1; otherwise, the value will be 0. The main disadvantage of this count vectorizer is that it cannot determine the importance of a specific word and cannot find linguistic similarity between words. The TF-IDF vectorizer overcomes the first disadvantage, which is determining the importance of the word. For this research, we will use the TF-IDF feature extraction technique. The TF IDF score is proportional to the importance of that word in the document. If the TF-IDF value is high, the importance of that particular word is also high. To calculate the TF-IDF value, first calculate the TF and IDF values. The final TF-IDF matrix is obtained by multiplying the TF and IDF values.

Term Frequency(TF) is the frequency of a word in a particular row(document). Tf is found using the formula,

$$TF(t) = c(t,d)$$

where,

c(t,d) - the number of times the term t appears in the document d

Following that is IDF, which stands for Inverse Document Frequency. This IDf value will be high if the word in a row (document) is uncommon in the other rows of that corpus. In IDF there is a consideration that if the occurrence of words in a specific document is reduced, that word will be more prominent. IDF is found using the formula

$$IDF(t) = 1+\log(N/df(t))$$

where,

N-the total number of documents in the corpus

Df(t)-number of documents containing the term t

Finally ,the TF-IDF value can be found using the formula,

$$TF\text{-}IDF(t) = TF(t)*IDF(t)$$

The Tf-IDF values vary depending on how many times the word appears within the row and how many times the word appears in the corpus. When the number of rows containing the same word increases, the TF-IDF value decreases. When the frequency of a specific word within that row increases, so does the TF-IDF value.

b) Random Forest Classifier: This is a Supervised Machine Learning algorithm that is primarily used for classification and regression tasks. The Random Forest Algorithm is composed of N individual decision trees. The model's output is predicted by combining the results of all the individual decision trees. The dataset is divided into subsets and distributed to each of the decision trees within the random forest classifier. These decision trees categorize the output for the specified subset. All of these individual tree results are taken into account when predicting the final output. The final result is calculated using a majority voting methodology.

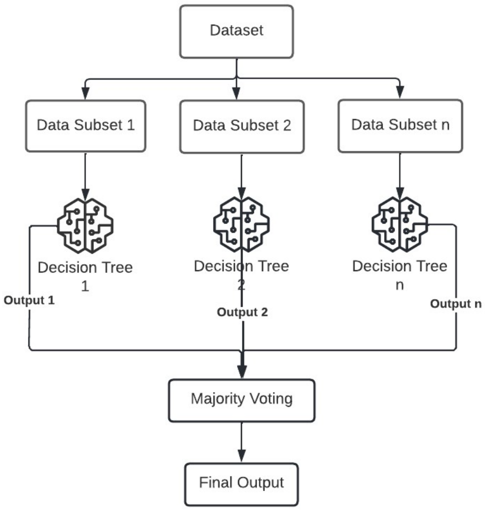


Fig. 2 Random Forest Classifier

The flow of the random forest classifier is depicted in figure 2. The dataset is initially divided into data subsets based on the number of decision trees considered. These data subsets are fed into the appropriate decision tree classifier, and the output is classified. These outputs from all decision trees are subjected to majority voting; here, the frequencies of the distinct classified outputs are calculated, and the class with the highest frequency is chosen and considered the final output.

B. Convolutional Neural Network based Tonal Analysis

Unlike the first method, this one uses deep learning for classification. Deep learning is a type of machine learning that employs neural networks. CNN is a deep learning method used for image analysis. As the CNN model does not train directly with audio data, it is converted into Mel spectrogram images and input into the model for training. The trained model is used to determine if the audio is depressed or not.

a) Convolutional Neural Network: During training, the data provided to CNN goes through many processes such as convolution, activation, and pooling. The input is convolved and passed to the next layer in the convolution process. Pooling is a process in which the dimensions of the input to the layer is based on the type of pooling defined like max pooling, min pooling, average pooling and global max pooling. The activation function allows neurons to be activated are not. Followed by other processes a fully connected neural network is established where classification is done. A dropout layer is declared to kill some neurons in the fully connected neural network so that the model does not overfit. The process flow of CNN is depicted in figure 3.

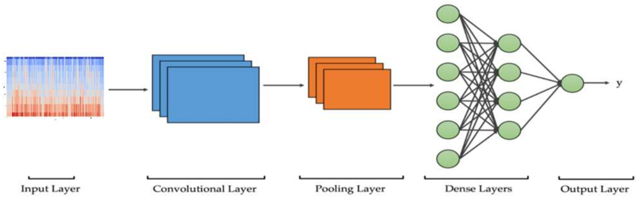


Fig. 3 CNN Process Flow

C. Ensembling Model

When the above two individual models are ensembled, the prediction accuracy can be gradually increased depending upon the ensembling methodology. The results of the CNN-based tonal analysis model and the Semantic analysis model are combined with the results of another machine learning model called Support Vector Machine (SVM).

a) Support Vector Machine(SVM): SVM is a supervised machine learning technique which can do both classification and regression. It is straightforward to use, which is one of its key advantages. The main goal of the model is to find the best hyperplane in multi dimensions to divide the classes during a classification problem. So that it may use the hyperplane to classify the new data points. if a new data point falls on the gap of a class created by the hyperplane , it belongs to that class.

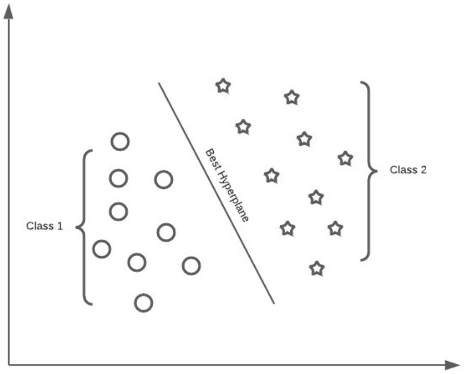


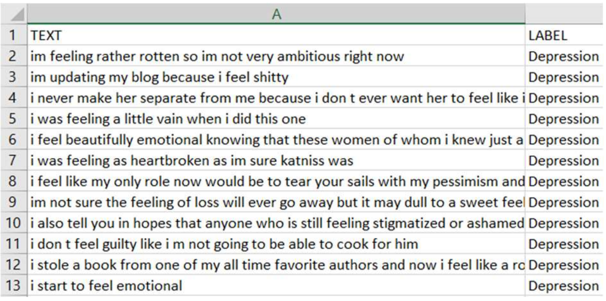
Fig. 4 Points plotted with SVM Algorithm

In the figure 4 data points belonging to two classes (class 1 and class 2) are plotted and a best hyperplane is established using SVM algorithm. If a new data point enters the space it will be mapped in either left or right gape of the hyperplane . Depending on the gape where the data point is present the class is identified.

4.MODULES

A. Data Collection

The most difficult aspect of this research is gathering the datasets. We need two types of datasets because we are implementing two different approaches for depression detection. The first is a textual dataset used to train the semantic-based detection model. The second dataset is an audio dataset that is used for a CNN-based approach to detecting depression. The textual depression data for the first approach was not easily accessible on the internet. The most commonly available data sets are twitter datasets, which are commonly used in this type of depression detection research. After conducting a thorough search for depression text data on the internet, we decided to create our own dataset using depressed keywords that were readily available.We found a dictionary with specific depression keywords in it and used those to frame sentences. We had created a text dataset and saved it in comma-separated values (CSV) format. The dataset contains 12500 rows, with nearly 6800 depressed sentences and nearly 5700 non-depressed sentences, which include both joyful and neutral dialogues. The depressed sentences are labeled "Depression," while the non-depressed sentences are labeled "No Depression."



	A	
1	TEXT	LABEL
2	im feeling rather rotten so im not very ambitious right now	Depression
3	im updating my blog because i feel shitty	Depression
4	i never make her separate from me because i don t ever want her to feel like i	Depression
5	i was feeling a little vain when i did this one	Depression
6	i feel beautifully emotional knowing that these women of whom i knew just a	Depression
7	i was feeling as heartbroken as im sure katniss was	Depression
8	i feel like my only role now would be to tear your sails with my pessimism and	Depression
9	im not sure the feeling of loss will ever go away but it may dull to a sweet feel	Depression
10	i also tell you in hopes that anyone who is still feeling stigmatized or ashamed	Depression
11	i don t feel guilty like i m not going to be able to cook for him	Depression
12	i stole a book from one of my all time favorite authors and now i feel like a rc	Depression
13	i start to feel emotional	Depression

Fig. 5 Textual Depression Dataset

Similarly, the audio dataset for the second approach is created manually. A total of 200 audios were recorded, Each audio file is recorded for 25 seconds on average. Among those 200 audio clips, 130 were depressed speech and 70 were non-depressed speech. The audio was recorded from people of various ages, both male and female. Initially, the audio was recorded in.ogg format.

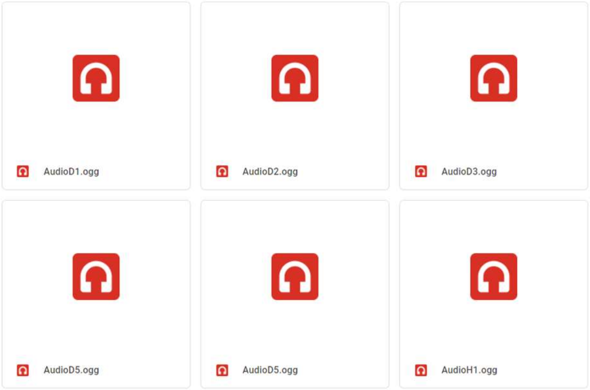
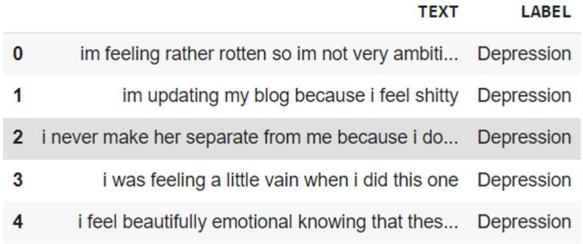


Fig. 6 Depression Audio Dataset

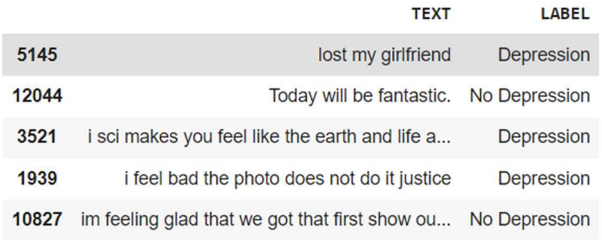
B. Dataset Preprocessing

The textual dataset, which has 12500 rows, is initially randomized for the semantic approach. As a result, the sentences for depression and no depression are mixed up before being divided into test and train datasets.



	TEXT	LABEL
0	im feeling rather rotten so im not very ambiti...	Depression
1	im updating my blog because i feel shitty	Depression
2	i never make her separate from me because i do...	Depression
3	i was feeling a little vain when i did this one	Depression
4	i feel beautifully emotional knowing that thes...	Depression

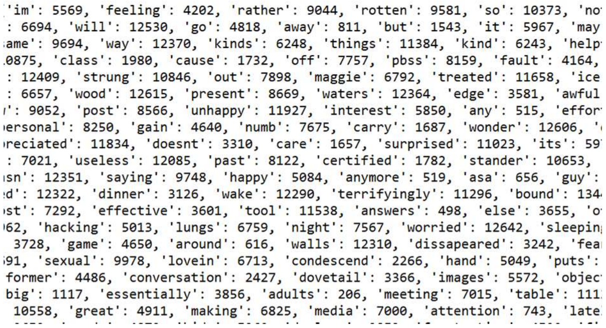
Fig. 7 Before Randomizing



	TEXT	LABEL
5145	lost my girlfriend	Depression
12044	Today will be fantastic.	No Depression
3521	i sci makes you feel like the earth and life a...	Depression
1939	i feel bad the photo does not do it justice	Depression
10827	im feeling glad that we got that first show ou...	No Depression

Fig. 8 After Randomizing

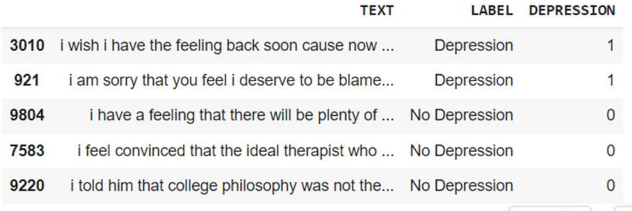
The dataset is divided into two columns: the first column contains sentences labeled "TEXT," and the second column contains corresponding labels. This "TEXT" column is fed into the tfidf vectorizer, which converts it to a matrix with TF-IDF features. Figure 9 shows the vocabularies created by the tfidf vectorizer.



```
'im': 5569, 'feeling': 4202, 'rather': 9044, 'rotten': 9581, 'so': 10373, 'no
: 6694, 'will': 12530, 'go': 4818, 'away': 811, 'but': 1543, 'it': 5967, 'may
ame': 9694, 'way': 12370, 'kinds': 6248, 'things': 11384, 'kind': 6243, 'help
0875, 'class': 1980, 'cause': 1732, 'off': 7757, 'pbss': 8159, 'fault': 4164,
: 12409, 'strung': 10846, 'out': 7898, 'maggie': 6792, 'treated': 11658, 'ice
: 6657, 'wood': 12615, 'present': 8669, 'waters': 12364, 'edge': 3581, 'awful
': 9052, 'post': 8566, 'unhappy': 11927, 'interest': 5850, 'any': 515, 'effor
ersonal': 8250, 'gain': 4640, 'numb': 7675, 'carry': 1687, 'wonder': 12606, '
reciated': 11834, 'doesnt': 3310, 'care': 1657, 'surprised': 11023, 'its': 59
: 7021, 'useless': 12085, 'past': 8122, 'certified': 1782, 'stander': 10653,
sn': 12351, 'saying': 9748, 'happy': 5084, 'anymore': 519, 'asa': 656, 'guy':
d': 12322, 'dinner': 3126, 'wake': 12290, 'terrifyingly': 11296, 'bound': 134
st': 7292, 'effective': 3601, 'tool': 11538, 'answers': 498, 'else': 3655, 'o
62, 'hacking': 5013, 'lungs': 6759, 'night': 7567, 'worried': 12642, 'sleepin
 3728, 'game': 4650, 'around': 616, 'walls': 12310, 'dissapeared': 3242, 'fea
91, 'sexual': 9978, 'lovein': 6713, 'condescend': 2266, 'hand': 5049, 'puts':
former': 4486, 'conversation': 2427, 'dovetail': 3366, 'images': 5572, 'objec
big': 1117, 'essentially': 3856, 'adults': 206, 'meeting': 7015, 'table': 111
 10558, 'great': 4911, 'making': 6825, 'media': 7000, 'attention': 743, 'late
```

Fig. 9 Vocabularies generated by tfidf vectorizer

The Depression and No Depression labels are further updated to 1 and 0 In which 1 denotes Depression and 0 denotes No Depression data.



	TEXT	LABEL	DEPRESSION
3010	i wish i have the feeling back soon cause now ...	Depression	1
921	i am sorry that you feel i deserve to be blame...	Depression	1
9804	i have a feeling that there will be plenty of ...	No Depression	0
7583	i feel convinced that the ideal therapist who ...	No Depression	0
9220	i told him that college philosophy was not the...	No Depression	0

Fig. 10 Labeling 1 And 0

The Audio data for CNN based tonal analysis is split into train and validation in the ratio of 3: 1 (150 audio for training and 50 audio for validation). The audio is converted to Mel spectrogram images for the second approach, i.e. the audio is converted into a visual representation so that image classification models such as Convolutional Neural Network can be used to analyze and identify the data. Mel spectrograms are used because they better represent audio spectral representation than other types of spectral representation of audio. Mel spectrograms represent audio signals in the same way that the human ear does. Mel spectrograms are simply spectrograms that have been converted to the Mel scale using nonlinear transformations.

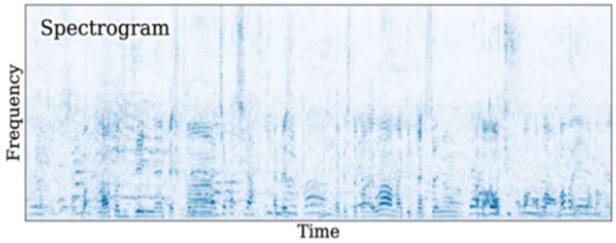


Fig. 11 Spectrogram Representation Of The Audio

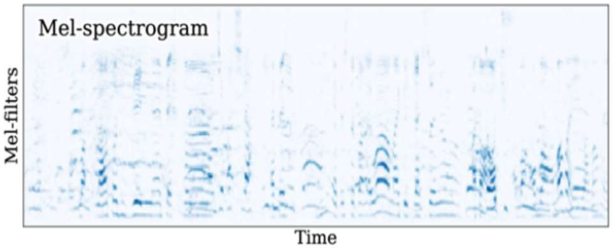


Fig. 12 Mel-Spectrogram Representation Of The Audio

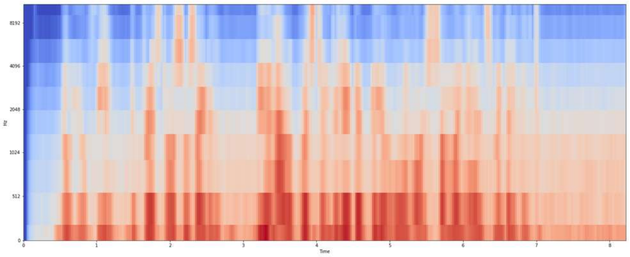


Fig. 13 Mel-Spectrogram After Encoding Color

C. Model Development

a) Semantic Analysis Model: The semantic analysis technique is used to proceed with the textual depression prediction portion. This is the process by which TF-IDF vectorisation is carried out. Initially, the 12500-row dataset is labeled as Depression and Non Depression. For the sentences in each row, preprocessing has been used. Stop words and newlines are removed, and lower case is used throughout the dataset. We randomized the rows before splitting the dataset into test and train to ensure a proper mix of depressed and non-depressed data rows. The results obtained will be more accurate as a result of randomization, and unwanted biases will be avoided. Following randomization, the dataset is split in a 3:1 ratio. In which 75% of the data is used for training and 25% for validation. The training dataset contains 51% depression text and 49% non-depression text, whereas the validation set is the opposite. The split dataset is fed into the model for training, which employs pipelining along with a classifier and vectorizer. The vectorizer is TF IDF, and the classification is done with the Random Forest algorithm. The number of feature parameters is varied, and the model is fitted and verified for accuracy. The feature value was initially set to 10000 and gradually increased to 20,000 and 30,000 to obtain the corresponding accuracies.

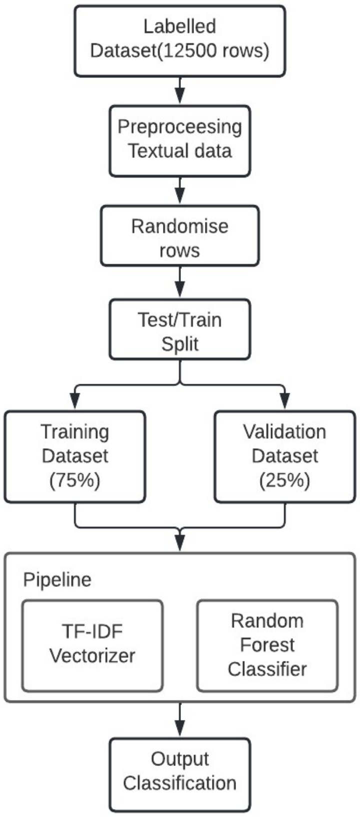
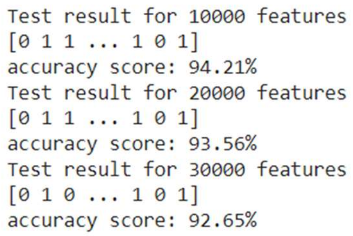


Fig. 14 Semantic Analysis Model

When it came to accuracy, the model fitted with 10000 features performed well for the given dataset, with a testing accuracy score of 94.21 percent.



```
Test result for 10000 features
[0 1 1 ... 1 0 1]
accuracy score: 94.21%
Test result for 20000 features
[0 1 1 ... 1 0 1]
accuracy score: 93.56%
Test result for 30000 features
[0 1 0 ... 1 0 1]
accuracy score: 92.65%
```

Fig. 15 Testing Accuracies Of Semantic Analysis Model

b) CNN based Tonal Analysis Model: The CNN is used for tonal analysis of a person's speech. The audio is converted into Mel spectrogram and fed into the CNN model for training.

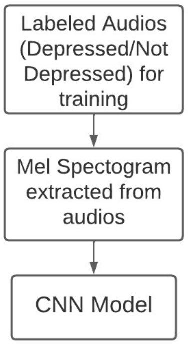

Labeled Audios (Depressed/Not Depressed) for training
Mel Spectogram extracted from audios
CNN Model

Fig. 16 CNN Model's Training Workflow

The CNN model consists of three convolution layers. Each convolution layer is followed by a Relu Activation layer and each activation layer is followed by a max pooling layer. Relu (rectified linear activation function) will only pass if the input is positive or it will pass 0 to the next layer. Relu is known for its simple and fast approach. The Maxpooling2d layer selects the maximum value in an input window. The window moves in strides along the data. Then the input value is passed into a fully connected network. The fully connected layer consists of two dense layers, one Relu activation layer and one Sigmoid activation layer. The sigmoid function takes the sum of input values and returns values ranging from 0 to 1. It is used when a probabilistic value must be returned as it only returns values between 0 to 1.

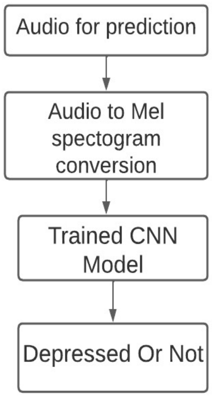
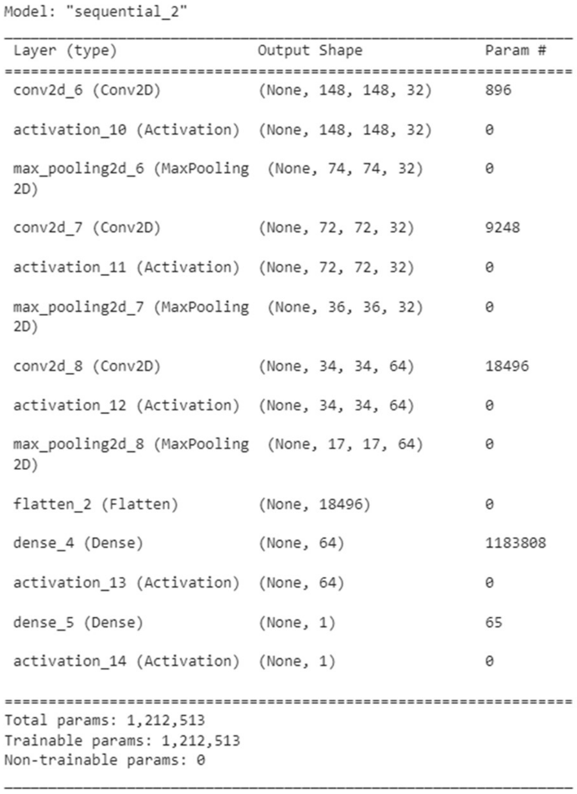


Fig.17 CNN Model's Testing Workflow

For prediction, the audio of a person's speech is converted into a Mel spectrogram and passed into the model for classifying if the person is depressed or not depressed.



```
Model: "sequential_2"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d_6 (Conv2D)           (None, 148, 148, 32)      896

 activation_10 (Activation)  (None, 148, 148, 32)      0

 max_pooling2d_6 (MaxPooling  (None, 74, 74, 32)       0
 2D)

 conv2d_7 (Conv2D)           (None, 72, 72, 32)        9248

 activation_11 (Activation)  (None, 72, 72, 32)        0

 max_pooling2d_7 (MaxPooling  (None, 36, 36, 32)       0
 2D)

 conv2d_8 (Conv2D)           (None, 34, 34, 64)        18496

 activation_12 (Activation)  (None, 34, 34, 64)        0

 max_pooling2d_8 (MaxPooling  (None, 17, 17, 64)       0
 2D)

 flatten_2 (Flatten)         (None, 18496)             0

 dense_4 (Dense)             (None, 64)                1183808

 activation_13 (Activation)  (None, 64)                0

 dense_5 (Dense)             (None, 1)                 65

 activation_14 (Activation)  (None, 1)                 0

=================================================================
Total params: 1,212,513
Trainable params: 1,212,513
Non-trainable params: 0
_________________________________________________________________
```

Fig.18 Model Summary

c)

d) Ensembling Model: Both approaches are used to classify 230 pieces of audio, and the results are saved in CSV format as 0's and 1's (0-non depressed and 1 - depressed). Both models' outputs are set as attributes, and the actual class of the audio is set as target and fed into an SVM for training. The SVM finds the best hyperplane with the greatest distance between two classes' data points. Figure xx clearly shows how Svm models are used to determine whether a person is depressed or not.

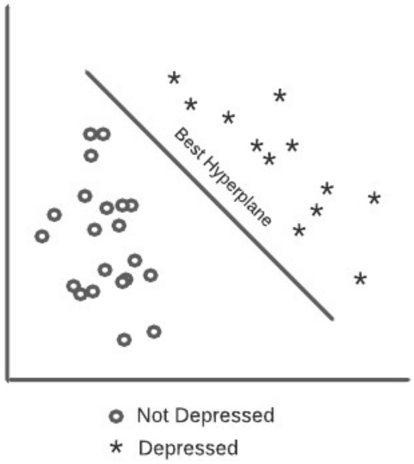


Fig. 19 Applying SVM to Depression dataset

D. Implementation

The person's audio is recorded and uploaded to the model during real-time prediction, and the text is extracted from the uploaded audio using Google Recognizer. This text is then passed to the pretrained semantic model on the one hand, and the person's input audio is passed directly to the tonal analysis model, which converts it to a MEL Spectrogram and predicts the output on the other. The real-time working flow of two distinct approaches is depicted in figure 19. Both models operate as a binary classification system, with the output being either 1 or 0 if depression is detected. The output of the two models is fed into an ensemble model, which combines the output of the tonal and semantic models to produce the final result. The mail is sent based on the final outcome. If the person is said to be depressed, counseling quotes as well as doctor contact information will be sent.So, the person can easily reach out to the consultant and get medications soon.

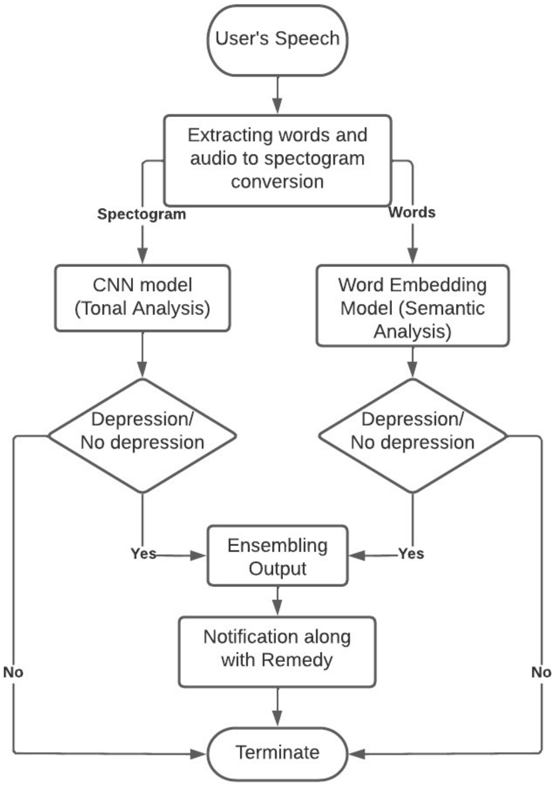
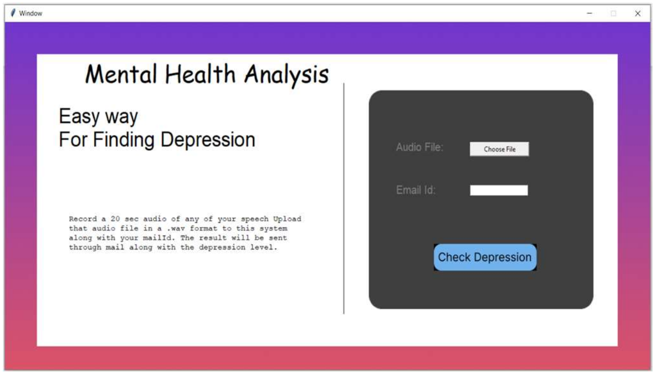

User's Speech
Extracting words and audio to spectogram conversion
Spectogram
Words
CNN model (Tonal Analysis)
Word Embedding Model (Semantic Analysis)
Depression/ No depression
Depression/ No depression
Yes
Yes
Ensembling Output
Notification along with Remedy
No
No
Terminate

Fig. 19 Realtime Depression Detection Workflow

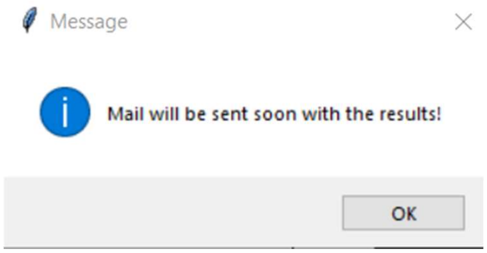
An user interface has been created using Python tkinter. The user must record the audio for approximately 25 seconds and upload it to the system in .wav format. The user must also provide a valid email address so that the result can be delivered to that address.



Record a 20 sec audio of any of your speech Upload that audio file in a .wav format to this system along with your mailId. The result will be sent through mail along with the depression level.

Fig. 20 Depression Detection UI

Following the completion of the details, on clicking on the button called "Check Depression", a notification message will be displayed, as shown in figure 21.

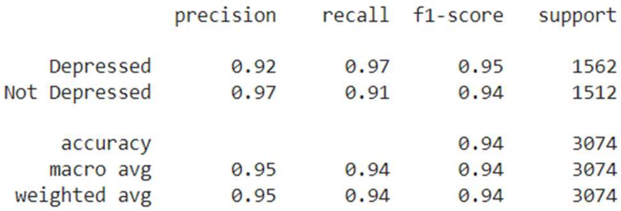


Mail will be sent soon with the results!

Fig . 21 Completion Message

5. EXPERIMENTAL RESULT AND DISCUSSION

With the implementation completed, the models are capable of accepting raw audio files and predicting whether or not the person is depressed. A Semantic analysis model and a Tonal analysis model have been successfully implemented and tuned for improved accuracy, and finally ensembling for these two models output with SVM classifier is done. Figure 22 depicts the metrics of the Semantic analysis models.



```
               precision    recall  f1-score   support

    Depressed       0.92      0.97      0.95      1562
Not Depressed       0.97      0.91      0.94      1512

     accuracy                           0.94      3074
    macro avg       0.95      0.94      0.94      3074
 weighted avg       0.95      0.94      0.94      3074
```

Fig. 22 Semantic Model's Classification Report

The training and validation accuracy of the CNN model is 0.8417 and 0.8667. The training and validation loss of the CNN model is 0.3657 and 0.3326. The training and testing accuracy of SVM classifier is 0.89 and 0.98

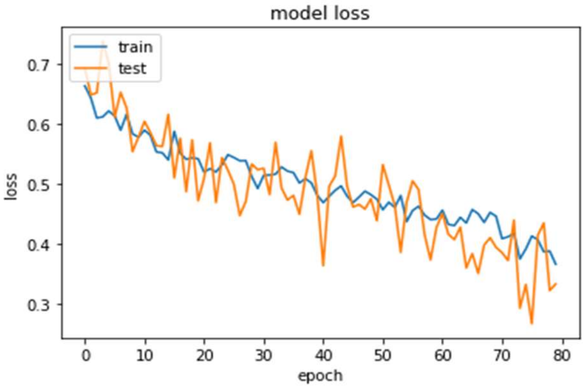


Fig. 23 CNN Model's Loss Graph

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

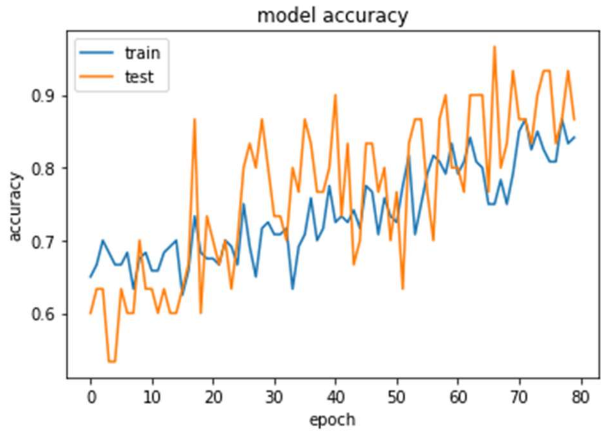


Fig. 24 CNN Model's Accuracy Graph

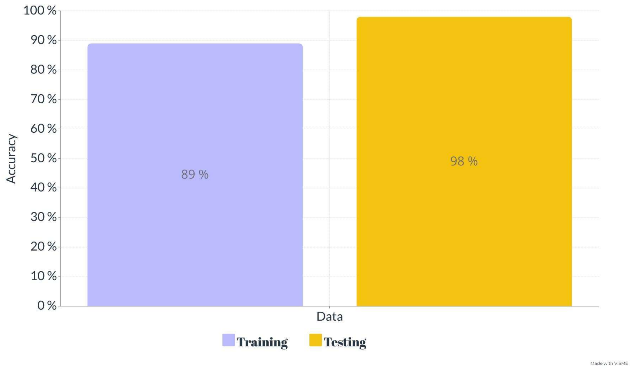
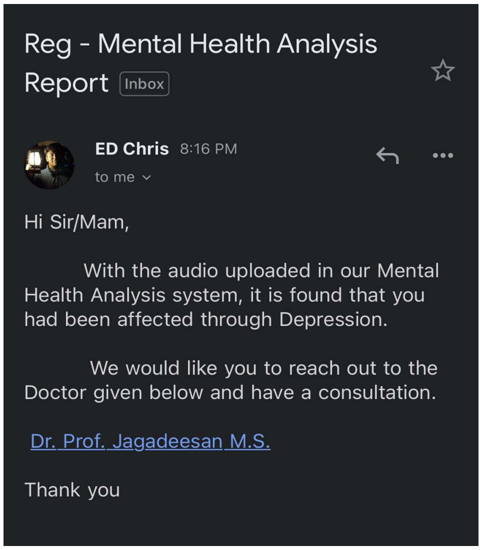


Fig. 25 Ensembling Model's Training and Testing Accuracy

If the model predicts that the person is depressed, it will send an email to that user with the content shown in figure 26. One of the best Doctors' contact information is also included in the email so that the user can easily schedule a consultation appointment.



Reg - Mental Health Analysis Report

ED Chris 8:16 PM
to me

Hi Sir/Mam,

With the audio uploaded in our Mental Health Analysis system, it is found that you had been affected through Depression.

We would like you to reach out to the Doctor given below and have a consultation.

Dr. Prof. Jagadeesan M.S.

Thank you

Fig. 26 Mail Notification with Prediction Result

GoogleMaps API has been used for finding the nearby Psychiatrist who is best in this treatment. With this system, depression can be identified in a very short period of time and with great precision. Immediate action could be taken to begin proper treatment in order to overcome it.

6. CONCLUSION

As a result, our research project would benefit many people who are not mentally healthy by detecting depression with a single piece of an audio file and also by providing information about a consultant for further treatment. With the help of this model, we can predict depression in minutes and at a much lower cost than other medical tests that are required to determine whether or not a person is affected by depression. It has the potential to save many people's lives by predicting depression at an early stage. Because this tool is easily accessible from the comfort of our own home.

DECLARATION:

Ethics Approval and Consent to Participate:

No participation of humans takes place in this implementation process

Human and Animal Rights:

No violation of Human and Animal Rights is involved.

Funding:

No funding is involved in this work.

CONFLICT OF INTEREST:

Conflict of Interest is not applicable in this work.

Authorship contributions:

There is no authorship contribution

Acknowledgement:

There is no acknowledgement involved in this work.

REFERENCES

[1] F.M.Shah, F. Ahmed, S.K.S Joy, S. Ahmed, S. Sadek, R. Shil and M.H.Kabir, "Early depression detection from social network using deep learning techniques," In 2020 IEEE Region 10 Symposium (TENSYMP) 2020,pp. 823-826.DOI: 10.1109/TENSYMP50017.2020.9231008.

[2] B.Parkar, S. Lanjekar, A. Mulla and V.Patil, " Depression Detection Using NLP Algorithm On Youtube Data," International Research Journal of Engineering and Technology (IRJET), vol. 8,2021, pp. 4, DOI:10.11591/eei.v12i2.4182.

[3] A. H. Orabi, P. Buddhitha, M.H. Orabi, D. Inkpen, "Deep Learning for Depression Detection of Twitter Users", In Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic, 2018,pp. 88–97, DOI:10.18653/v1/W18-0609.

[4] F. Sadeque, D. Xu, and S. Bethard, "Measuring the Latency of Depression Detection in Social Media", Eleventh ACM International Conference on Web Search and Data Mining, New York, NY, USA, 2018,pp. 495–503. DOI:10.1145/3159652.3159725.

[5] C. Lin, P. Hu, H. Su, S. Li, J. Mei, J. Zhou, and H. Leung, "SenseMood: Depression Detection on Social Media", 2020 International Conference on Multimedia Retrieval, New York, NY, USA,2020, pp. 407–411. DOI:10.1145/3372278.3391932.

[6] Y. Gong and C. Poellabauer, "Topic Modeling Based Multi-modal Depression Detection", 7th Annual Workshop on Audio/Visual Emotion Challenge, New York, NY, USA, 2017,pp. 69–76, DOI:10.1145/3133944.3133945.

[7] G. Shen, J. Jia, L.Nie, F. Feng, C. Zhang, T. Hu, T.S.Chua and W. Zhu,"Depression Detection via Harvesting Social Media: A Multimodal Dictionary Learning Solution", Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17,2017,pp. 3838—3844. DOI:10.24963/ijcai.2017/536.

[8] M. M. Tadesse, H. Lin, B. Xu and L. Yang, "Detection of Depression-Related Posts in Reddit Social Media Forum," IEEE Access, vol. 7,2019, pp. 44883-44893. DOI:10.1109/ACCESS.2019.2909180.

[9]Q.Zhou ,J. Shan ,J. Ding,"Cough Recognition Based on Mel-Spectrogram and Convolutional Neural Network", Front Robot AI, vol. 8 2021,PP.580080. DOI:10.3389/frobt.2021.580080.

[10] H. Meng, T. Yan, F. Yuan and H. Wei, "Speech Emotion Recognition From 3D Log-Mel Spectrograms With Deep Learning Network," in IEEE Access, vol. 7, 2019,pp. 125868-125881. DOI:10.1109/ACCESS.2019.2938007.

[11]K.V.Kumar and A.Rajaram, "Energy efficient and node mobility based data replication algorithm for MANET," International Journal of Computer Science, 2019.

[12]A.P.Sridevi and A.Rajaram, "Efficient Energy Based Multipath Cluster Routing Protocol For Wireless Sensor Networks". Journal of Theoretical & Applied Information Technology,vol.68,2014.

[13]A.Rajaram and S.Kannan,"ENERGY BASED ROUTING ALGORITHM FOR MOBILE AD HOC NETWORKS," Journal of Theoretical & Applied Information Technology, Vol.61, 2014. **DOI:** 10.1109/WD.2008.4812884

[14]A.Rajaram and J.Sugesh, "Power aware

routing for MANET using on-demand multipath routing protocol," International Journal of Computer Science Issues (IJCSI), vol.8, 2011,pp.517. DOI:10.1109/ICDCSW.2004.1284112

[15]A.Rajaram, and A.Baskar, "Hybrid Optimization-Based Multi-Path Routing for Dynamic Cluster-Based MANET," Cybernetics and Systems.2023. https://doi.org/10.1080/01969722.2023.2166249