International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
doi : https://doi.org/10.32628/IJSRST229437

Stress Detection Through Speech Analysis Using Machine Learning

Vamsinath J[1], Varshini Bonagiri[2], Sandeep T[2], Meghana V[2], Latha B[2]
*[1]Computer Science and Engineering (Data Science), VNR VJIET, Hyderabad, Telangana, India
[2]Computer Science and Engineering, VNR VJIET, Hyderabad, Telangana, India

Article Info
Volume 9, Issue 4
Page Number : 334-342

Publication Issue
July-August 2022

Article History
Accepted : 10 July 2022
Published : 22 July 2022

ABSTRACT

Stress is a common issue for every person. Around 264 million individuals suffer from stress, which is one of the most frequent psychological issues. We present a deep learning-based stress detector model using audio signals. The main aim is to differentiate stressed and non-stressed speeches. The deep learning algorithm used here is Convolutional Neural Network(CNN) which is made up of connected layers that are all related.Speech is transformed into spectrograms and they are fed to the Convolutional Neural Network (CNN) model. Features are extracted using Mel-frequency cepstral coefficients from pre-processed data.

Keywords : Convolutional Neural Network, Mel-frequency cepstral coefficients, Librosa

I. INTRODUCTION

Stress is a normal reaction for any human being, their reaction has both positive and negative effects in everyone's life. Automatic stress detection is becoming a frequently investigated problem in human-computer interface models, as the demand for communication between humans and intelligent systems grows. So, while measuring hormone levels (e.g., cortisol) can detect stress, it is not a practical way for detecting stress in human-machine interactions..

More than 264 million individuals suffer from stress, which is one of the most frequent psychological issues. Major stress can lead to suicide and substance abuse. Detecting and treating stress is a top priority. The traditional approach to diagnosing stress is conducting clinical interviews to examine the mental status of the person.The purpose of this document is to provide you with some guidelines. You are, however, encouraged to consult additional resources that assist you in writing a professional technical paper.

It proposes the Convolutional Neural Networks model to classify stress using speech features in the RAVDESS dataset. Speech is processed in three steps in this project: pre-processing,classification and feature extraction. In the pre-processing stage the unwanted voices and disturbances are removed from the speech. In Feature extraction stage features are extracted from the processed data then it is converted into spectrograms and fed as input to the CNN model. Then the model predicts the output whether the person is stressed or unstressed.

The levels of particular hormones like cortisol are being used to consistently detect stress. In Fact the aim of this project is to automate the process of stress detection without the intervention of a Doctor or Psychiatrist. This project proposes a hybrid deep learning model to analyze whether the person is stressed or unstressed using speech.

II. PROPOSED SYSTEM

We present a deep learning-based psychological stress detection method, on the basis of voice data. Automatic stress detection is becoming an interesting research subject as the demand for communication between human-to-human and human-to-intelligent-systems grows. The amount of particular hormones (e.g., cortisol) can be used to consistently detect stress. The suggested technique extracts Mel Filter bank coefficients from pre- processed speech data and then uses CNN and dense fully connected neural networks to estimate the state of stress output(i.e., stressed or unstressed). In feature extraction the audio file is converted to spectrogram, because sound or signals vary with time. Spectrogram is fed to the model to predict whether the person is stressed, unstressed or neutral.

III. RESULTS AND DISCUSSION

A. Literature Survey

Nikolaos [1] used Berlin Emotional Database (EMO-DB) datasets for emotion detection from speech data. Using Praat, 133 speech features are extracted from the pitch and MFCC's, as well as energy and forms from the speech waveform, then analyzed to build a feature set. A person's emotion is divided into seven categories, which are further divided into two hyper categories: high and low arousal emotion. The classifier recognises high arousal emotion 100% of the time and low arousal emotion 87% of the time. WEKA and SVM are the classifiers applied here.

Maheshwari [2] observed speech emotion recognition incorporates spectral and prosodic characteristics. MFCC is a spectral feature and it contains four operations on human speech like fast fourier transform, Mel filter bank, logarithmic conversion and finally discrete cosine transform. Radial basis and back propagation are used to recognize the emotions based on the selected frames. Radial basis function network recognizes human emotions more accurately. Dhas [3] choose RAVDESS dataset for stress detection from speech. The suggested module is made up of eight CNN layers and entirely connected layers. Using Mel- frequency cepstral coefficients with pre-processed input, feature extraction is performed, and the output is predicted using a binary decision criterion using a Convolutional Neural Network. Filtering, windowing and framing are a few examples of pre-processing phase. The pre-emphasis filter is used to increase the spoken signal's intensity in higher frequencies.

Nisha [4] presents that stress can be detected using image processing. Theano framework is used to improve the execution and development time of the Linear regression model, is used as a deep learning algorithm, then it builds a model for predicting stress. Here eyebrow movement plays an important role because stress is detected if a person shows high variations of eyebrow movement constantly in fixed intervals. The Deep Learning model after training predicts whether the person is stressed or not.

Tanish [5] observed that DAIC-WOZ contains a PHQ8 score of patients; this dataset is used for depression analysis using audio features. Audios are transformed into spectrograms using Librosa and spectrograms are fed to CNN model with average pooling layers to know if the patient is depressed or non-depressed. Following feature extraction, each audio file is fed into a 6-layer convolutional neural network with a 32-filter input layer and a 3x3 kernel size, followed by an activation function called ReLU. Flatten layer is used as a second layer in CNN that is used for categorization and has a total of 100 epochs. Overall accuracy is 81% and output is in binary format.

Hea [6] used AVEC2013 and AVEC2014 datasets combined into a single database with 340 video clips of 292 people to predict depression scale. They manually extract Low Level Descriptors (LLD) from unprocessed voice recordings and state-of-the-art texture characteristics known as Median Robust extended Local Binary Patterns (MRELBP) from audio spectrograms. Second, the Deep Convolutional Neural Network(DCNN) is designed to extract features from spectrograms and unprocessed voice waveforms. They used Median Robust Extended Local Binary Patterns (MRELBP) on a spectrogram to extract textural data, and then they used the openSMILE tools to extract audio features. For depression scale prediction, firstly we compare the performance of the LBP feature with that of the MRELBP feature, and then compare the performance of hand-crafted characteristics with that of deep-learned characteristics.

Savita [7] used eleven subjects to be interviewed and their responses are recorded on questions from two domains. Two men, three teenagers and two women are considered as first-domain and they are questioned about a previous situation in which they were humiliated. In the second domain one man and three women from housekeeping staff are asked about a stolen mobile and recorded the voice .The recordings were analyzed using PRAAT software. Spectrograms are drawn between normal speech and stressed speech. The datasets used are very less in the project.

Lee [8] defines for emotion recognition Interactive Emotional Dyadic Motion Capture database is used for extracting deep learning features using two different faced visual cues. This convolutional 3D network is used to detect and analyze the depression level.RNN is then used to get the further Depression levels.The RNN with a sequence of C3D features as input, improves the results. Because of functions like maxpool, a Convolutional neural network is substantially slower. CNN contains multiple layers, the training process will take more time if the machine does not contain a powerful GPU.

Adnan [9] observed EEG (Electroencephalography) signals sensors provide great accuracy, however they are not always available due to the rigorous conditions of their application. GSR, ECG, EMG, ST, and RESPIRATION sensors or the combinations of them, are used in the second mode. It is simpler to use these sensors than it is to use EEG. Following the acquisition of signals from these sensors. We can process signals using the Wavelet transform.

Nandita [10] mainly focused on the people where they monitor their stress level on a daily basis, such as vehicle drivers and sick patients etc.. Because stress cannot be quantified directly, we must use primary measures to extract specific properties. Isolated or combined metrics are used to identify them. They collected physiological and physical signals to determine stress. Genetic Programming is also used to determine emotions. Also generate regression equations from facial expression to measure emotions. Computational techniques can be used to find the best sensor fusion and automate data processing for stress detection and categorization. The study finishes with a summary and recommendations for further research on computational stress.

Ramesh [11] used the USC-IEMOCAP database to recognize emotion using speech features and transcriptions. To identify emotion, the CNN model described in this section uses speech transcriptions in the form of word embeddings as input. Without any prior knowledge of their semantic contexts, CNNs can be applied directly to word embeddings. The following model employs a Mel-frequency Spectrogram as input to a 2D CNN, and then applies a set of four parallel 2D convolutions to the

Spectrogram to extract its features, along with a CNN model for emotion detection. Overall accuracy of the model is 76.1%.

Stefan [12] presented how to find the stress level in the person. In this experiment, fifteen subjects were instructed to play a game in which they had to control an air simulation that became increasingly difficult to control as the game progressed. The participants were asked questions during the game, which were subsequently labeled by another group of subjects in order to acquire an exact target value for each of the replies. After training, an RNN is utilized to detect the degree of stress in the utterances. At a rate of 25 Hz, the neural network calculated the level of stress. Difficulty levels of the air controller simulation are Level, Planes, Speed , Target Size , Duration. By Using more data sets we can predict better automatic stress detection. The labeled echo state network (ESN) was developed and trained using labeled speech data. By this ESN we can predict the stress.

David [13] built a mobile application using voice features for stress monitoring, the software is simple to use and does not require the insertion of any equipment into the body. It's a web interface for a server application. Cardiac rhythm, skin temperature, increased skin electrodermal activity, and changes in body postures and speech patterns are all signs of stress. This programme uses just features extracted from voice to detect stress and sends the findings to a server application for telemonitoring. Mel frequency cepstrum coefficients, delta and delta-delta coefficients, and pitch are the voice features utilized by the device (Voice signals have multiple frequencies and constantly changes in time, pitch can be estimated through Pitch estimators). The experimental data for this is collected from the "Universal stress model" generated from the "Trier social stress test(TSST)" experiment; it employs speech signals recorded as ".wav" files, which are processed on a computer using a Java application. SVM (Support vector machine) is used as a classifier in this model. Creating telemonitoring apps entails both client and server application development. It has a 78 percent accuracy rate and might be enhanced by adding a speaker recognition module. The server application enables mobile apps to be used for telemonitoring in a variety of circumstances, including: monitoring the state of people who require daily support, detecting stressful behaviors, and monitoring patients by a psychiatrist.

Iain [14] observed that many factors contribute to an Emotion Phenomenon. Scientists such as Izard, Charles Darwin, Oatley, and Scherer explained the concept of emotions. Emotions differ from moods in that emotions occur rapidly and last seconds or minutes, but moods are more ambiguous in origin and might endure hours or days. The current physiological and psychological state of a person depends on the Current external stimuli and previous memory stimuli. These two factors are responsible for the outwardly perceivable emotion shown by the person. Scherer (1984) noted that emotions can be visualized in two or at most, 3D space which can be characterized as follows. (i)"Strength" (ranging from contempt to fear and surprise) (ii) "Valence'' corresponds to pleasantness-unpleasantness (ranging from love and happiness to anger) and (iii) "Activity" ranges from sleep to tension. Emotion is a nuanced blend of the three main levels of audio abstraction (suprasegmental, segmental, and intra segmental) turally distinct. This data is used to create a set of rules for controlling a high-quality voice synthesizer in order to generate simulated emotion effects in the output speech. High-quality synthesizers with appropriate levels of user control are now commercially accessible, and text-to-speech systems with emotion effects are technically viable. Because emotion is such a significant aspect of human speech, it will almost certainly be included in commercial speech synthesis systems.

Mikael [15] observed that the concepts of recurrent neural networks, RNN can be based on previous inputs, back propagation learning is characterized with feed forward networks, adapted to suit our modeling needs and extended to cover recurrent networks. The mapping process for a two-layered network consists of two steps i.e., y(t) = G(F(x(t)) (back propagation is used to discover the network's weights, which are G and F. RNNs differ from feed forward architecture in that they function not only on an input state but also on an internal state space - a record of what the network has already processed. This is the Iterated Function System's counterpart (ITF). According to y(t) = G(s(t)) and s(t) = F(s(t-1), the state space allows for the depiction of incrementally expanded dependence over unspecified intervals (t). When the desired output pattern exists, any network may be trained via back propagation. Back propagation, like conventional gradient descent, operates by computing the gradient of a cost(or error) function with respect to each changeable weight and then altering it accordingly. Summed Squared Error is the most commonly used cost function (SSE). The arrangement of the state space simply reflects the component elements of the training data for simpler tasks (learning grammar created by small finite state machines). The extremely non-linear, continuous space offers different sorts of dynamics for more demanding tasks (e.g., where a longer memory trail is required and context dependency is visible). Hierarchical Cluster Analysis (HCA) and eigen value and eigen vector characterizations are some of the analysis methodologies used.

Dhruvi desai [16] Fundamental frequency, ener signal, zero crossing rate(ZCR), LPC(linear predictor coefficient), MFCC, DTW (Dynamic Time Warping), SVM (Support Vector Machine), K-NN(K-Nearest Neighbor) and GMM (Gaussian Mixture Model) are some of the speech features included. It requires natural man-machine interaction. The feature extraction like MFCC gives the best accuracy on all the databases provided using spectral coefficients and linear kernel extracted from LPC(Linear Predictive Coding), classifier like ANN, GMM, KNN, SVM, Hidden Morkov Model(HMM) etc. The following is how a speech emotion identification system works: speech input is given, features are extracted, features are selected, and finally a classifier is used to recognize emotion. The phases in the feature extraction technique are as follows: To obtain Mel Spectrum, pre-emphasis, framing, windowing, rapid fourier transform, mil filter bank, and discrete cosine transform are used to the input (emotional speech). They used DTW, ANN, SVM, KNN, Gaussian Mixture Model(GMM). It has taken Berlin Emotion Database and Neural Emotion Database. Speech Emotion Recognition based on a variety of speech characteristics and classifiers is illustrated. MFCC achieves great recognition accuracy by providing a high level of perception of voice. DTW technique's accuracy (using Euclidean distance) was 57.5%. When compared to other classifiers, the accuracy of ANN is rather poor. KNN classifiers have a quick computation speed, making them a viable option if time is a factor. When the amount of speech features in the training phase rose, GMM's computational time increased.

Woodrow [17] observed for some pupils, it identifies a problem in language learning and has a crippling influence on their ability to speak English. Teachers should assist students in reducing their concern about learning a second language. Students studying English in Australia were found to benefit from a dual conceptualization of second language speaking anxiety as indicated by the second language speaking anxiety scale. It is utilized with lower-level second language speakers, as well as students who have begun university courses, to examine contextual influences on second language speaking anxiety. This study backed up the idea of anxiety stages, stating that a student may experience anxiety as a result of a skills shortage, which can be applied in the classroom. Skills deficient pupils can benefit from desensitization and relaxation treatments. Communication is required both within and outside the classroom, which can be accomplished by assigning out-of-class projects that make use of the vast language resources available.

Hui-ju [18] aimed to look at the language anxiety and learning motivation of 155 EFL students from the same Taiwanese private high school, 60 of whom were on the academic track and 95 on the vocational track. Their findings revealed that both groups experienced moderate degrees of language anxiety, with academic track students having higher levels of eccentric motivation and total learning motivation than vocational track students. Further ,it was discovered that there is a negative link between motivation and anxiety. This research explained motivation in language learning , foreign language anxiety and some research questions which were asked to students. Academic-track students have high extrinsic motivation when compared to vocational-track students. As a result, instructors should devote more time to helping students develop extrinsic incentive and boost intrinsic motivation, both of which are necessary for students to learn English more efficiently. As two groups of students experienced a mild level of anxiousness while learning English, teacher-student relationship should be built creating a more relaxing and lighthearted surroundings in the classroom.

Balli [19] describes the data of an individual stress through the data collected from smartphones touch screen panel, accelerometer and Gyroscope sensors through the writing behavior on the smartphone keypad. The initial stress detection is investigated using smartphone motion sensors and keyboard usage behavior, as well as Machine Learning approaches. The accuracy of categorization was assessed using cross validation and gain ratio feature selection procedures. This stress detection application can be implemented in a variety of ways, including using other smartphone motion and position sensors, implementing new and effective feature extraction algorithms, using devices with various internal sensors, such as smartwatches, and extracting more efficient feature subsets using dimension reduction and feature selection algorithms.

Chengwei [20] used five main sections: They used an online learning framework to investigate the SER system, moving from acted data to more naturalistic data. Our performed data and inquired data are described in Section 2. It performs an acoustic study of emotional characteristics. They discussed the online learning process and speech emotion recognizer in this article. Finally, they reported experimental results that suggest that utilizing online learning, they may combine action and elicited data. The challenge with online learning algorithms is combining existing offline classifiers and applying model parameters to a small amount of fresh online data quickly.

Kevin [21] summarized the test results of a candidate in an interview in the HR recruitment process through several speech analyses that have been conducted to detect the stress level of the interviewed candidate. Automatic video analysis is adopted in an HR screening product which automatically detects stress during an interview. Screening process has two phases. In the first phase the candidate receives an invitation for screening. In the second phase the candidate connects through the website and is asked a few questions. The candidates data records were analyzed and validated. To increase the accuracy score, features such as formants, MFCCs deltas or speech rate could be added.

Sharma [22] introduced Multiband Demodulation Analysis, where the audio signal is processed through a bank of fixed bandpass filters simultaneously, resulting in various time domain signals from speech signals. The Hilbert Transform or the TEO are then used to depict these signals in terms of their amplitudes and instantaneous frequencies. Though the sinusoidal and AM-FM representations of speech are viable alternatives to traditional speech analysis and processing, they do have significant drawbacks.

Himashu [23] proposed a deep learning approach for depression classification using audio features in the DAIC-WOZ dataset.In this, stress detection classification is carried out in three stages i.e., preprocessing, classification and feature extraction. Unwanted silence and the interviewers are removed from the audio signal and the preprocessing stage. The audio characteristics are extracted using the given distress analysis interview Corpus-Wizard of Oz(DAIC- WOZ) dataset, and a spectrogram is created, which is then fed to the neutral network model during the feature extraction stage. The method needs to be optimized to a greater percentage of efficiency.

Bhanusree [24] used acoustic features for training a classification model to categorize whether humans are not-depressed or depressed. The classifiers are trained using the DIAZ-WOZ database. COVAREP toolkit is used to extract prosodic, spectral, and voice control information. SMOTE analysis is used for overcoming the class imbalance. The digital data is processed, and depression-related characteristics are identified, followed by PCA (Principal Component Analysis) to minimize feature space dimension. SVM, Logistic Regression, and Random Forest were three machine learning techniques used in this study.

B. Drawbacks of Existing System

- The extraction of elements from speech signals in order to describe a person's emotional state is a serious problem.
- In the existing system the noise and echo are detected so that sometimes stress of the person can't be detected.
- It gives less accuracy and takes more time.
- Dataset used for the existing system is small.

C. Figures and Tables

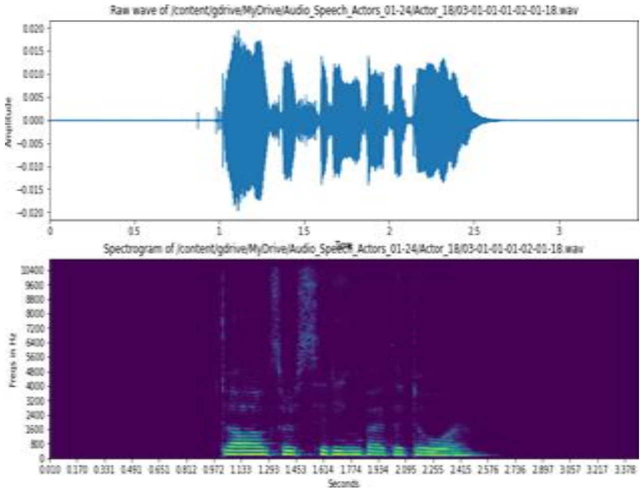

Figure 1: Conversion of audio file to spectrogram

	actualvalues	predictedvalues
40	male_stressed	male_stressed
41	male_stressed	male_stressed
42	male_unstressed	male_neutral
43	male_stressed	male_stressed
44	male_stressed	male_stressed
45	male_stressed	male_stressed
46	male_stressed	male_stressed
47	male_unstressed	male_neutral
48	male_stressed	male_stressed
49	male_unstressed	male_neutral
50	male_stressed	male_stressed
51	male_stressed	male_stressed

predictedvalues	actualvalues
male_neutral	262
male_stressed	244
male_unstressed	94

Figure 2 : Comparison of Actual value and Predicted Values

actualvalues	predictedvalues
male_neutral	120
male_stressed	320
male_unstressed	160

Figure 3 : Count of Predicted Values and Actual Values

D. Result

We have tested the algorithm with a few audio files with different age, gender. The following are the results of the testing shown in Table 1.

Parameter	**Value**
Accuracy	76.0
f1 Score	74.07526145118014

Table 2: Accuracy and f1 score

IV. CONCLUSION

The main purpose of this project is to get a better insight of the trends in stress detection techniques. For a better analysis, look back over the last few years to observe how it has changed and continues to change. This project deals with how to detect stress using deep learning techniques (i.e., CNN) using speech. In this process, feature extraction and selection of a model play a crucial role in stress detection.

V. FUTURE SCOPE

Interaction of human beings with the devices will be effortless incorporating this system in any electronic devices. This model highlights the use of speech data for stress detection; however, we may expand the model by working with text, image, and video data. It will take some time to optimize and generalize the model in order to attain high accuracy. We have only worked with English accent data so far, but we can deal with additional languages of audio data in the future.

VI. REFERENCES

[1]. Christos-Nikolaos, Anagnostopoulos and Theodoros Ilioua: “Towards Emotion Recognition from Speech: Definition, Problems and the Materials of Research” in March 2010.

[2]. Maheshwari Selvaraj, Dr.R.Bhuvana, S.Padmaja: “ Human Speech Emotion Recognition” in IJET, Vol 8 No 1| Feb-Mar 2016.

[3]. Dr. S. Vaikole, S. Mulajkar, A. More, P. Jayaswal, S. Dhas: “Stress Detection through Speech Analysis using Machine Learning” in (IJCRT) Volume 8, Issue | 5 May 2020.

[4]. Nisha Raichur, Nidhi Lonakadi, Priyanka Mural: “ Detection of Stress Using Image Processing and Machine Learning Techniques ” International Journal of Engineering and Technology (IJET), Vol 9 No 3S | July 2017.

[5]. Himani Negi, Tanish Bhola, Manu S Pillai , Deepika Kumar: “A Novel Approach for Depression Detection using Audio Sentiment Analysis” in Proceedings of 4th International Conference on Computers & Management (ICCM)2018.

[6]. Lang Hea, Cui Caob: “Automated depression analysis using convolutional neural networks from speech” in Journal of Biomedical Informatics 83 (2018) 103–111.

[7]. Savita Sondhi,Munna Khan,Ritu Vijay,Ashok K. Salhan: “Vocal Indicators of Emotional Stress”International Journal of Computer Applications (0975 – 8887) Volume 122 – No.15, July 2015.

[8]. Jinkyu Lee, Ivan Tashe: ,”High-level Feature Representation using Recurrent Neural Network for Speech Emotion Recognition” September 2015.

[9]. Bhanusree Yalamchini,Nikhil Sai Kota, Maruthi Saketh Abbaraju, Venkat Sai Sathwik Nadella, Sandeep Varma Alluri "Real-time Acoustic based Depression Detection using Machine Learning Techniques(ic-PETITE)", 2020.

[10]. Himashu Churi, Parul Keshri, Simran Khamkar, Prof. Amruta Sankhe "IRJET-A Deep Learning Approach for Depression Classification using Audit Features".Issue | 3 March 2021.

[11]. Rajib Sharma,LeandroVignolo, Gaston Sclotthauer,M.A. Colominas, H.Leanardo Rufiner,

S.R.M Prasanna "Empirical Mode Decomposition for Adaptive AM-FM Analysis of Speech '' VOL.88,pp,39-64,2017.

[12]. Adnan Ghaderi, Javad Frounchi,Alireza Farnam: “Machine Learning-based Signal Processing Using Physiological Signals for Stress Detection” 22nd Iranian Conference on Biomedical Engineering(ICBME 2015), Iranian Research Organization for Science and Technology (IROST), Tehran, Iran, 25-27 November 2015.

[13]. Nandita Sharma, Tom Gedeon: “Objective measures, sensors and computational techniques for stress recognition and classification” Information and Human Centered Computing Research Group, Research School of Computer Science, Australian National University, Canberra, ACT 0200, Australia.

[14]. Suraj Tripathi, Abhay Kumar, Abhiram Ramesh, Chirag Singh, Promod Yenigalla: “Deep Learning based Emotion Recognition System Using Speech Features and Transcriptions” in 2019.

[15]. Stefan Scherer, Hansjorg Hofmann ,Malte Lampmann,Martin Pfeil,Steffen Rhinow, Friedhelm Schwenker, Gunther Palm: “Emotion Recognition from Speech: Stress Experiment.” Institute of Neural Information Processing, Institute of Information Technology,University of Ulm, 89069 Ulm, Germany.

[16]. Virginia Sandulescu, Sally Andrews, David Ellis, Radu Dobrescu, Oscar Martinez-Mozos: “Mobile App for Stress Monitoring using Voice Features”, The 5th IEEE International Conference on E-Health and Bioengineering - EHB 2015.

[17]. Iain R. Murray and John L. Arnott: “Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion” The MicroCentre, Department of Mathematics and Computer Science, The University, Dundee DD] 4HN, Scotland, United Kingdom(Received 16 November 1990; revised 22 October 1992; accepted 23 October 1992

[18]. Mikael Bod: “A guide to recurrent neural networks and backpropagation” School of Information Science, Computer and Electrical Engineering Halmstad University. November 13, 2001.

[19]. Dhruvi desai: “Emotion recognition using Speech Signal: A Review” ME student, Communication System Engineering (E&C), Sarvajanik College of Engineering & Technology Surat, Gujarat, India.

[20]. Lindy Woodrow: “Anxiety and Speaking English as a Second Language” Article in RELC Journal · December 2006.

[21]. Hui-ju Liu1 & Chien-wei Chen2: “A Comparative Study of Foreign Language Anxiety and Motivation of Academic- and Vocational-Track High School Students”, Online Published: February 13, 2015

[22]. Ensar Arif Sagbas, Serkan Balli,Serdar Korukoglu: “Stress Detection via Keyboard Typing Behaviors by using Smartphone Sensors and Machine Learning Techniques.”, February, 2020

[23]. Kevin Tomba, Joel Dumoulin, Elena Mugellini, Omar Abou Khaled,Salah Hawila: “Stress Detection Through Speech Analysis”, 2018.

[24]. Chengwei Huang, Ruiyu Liang, Qingyun Wang, Ji Xi,Cheng Zha, Li Zhao: “Practical Speech Emotion Recognition Based on Online Learning:From Acted Data to Elicited Data”, 4 June, 2013.

Cite this article as :

J. Vamsina Vamsinath J, Varshini Bonagiri, Sandeep T, Meghana V, Latha B, "Stress Detection Through Speech Analysis Using Machine Learning", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 9 Issue 4, pp. 334-342, July-August 2022. Available at doi : https://doi.org/10.32628/IJSRST229437
Journal URL : https://ijsrst.com/IJSRST229437