Detection of Human Mental Stress Using Speech Signals

Mrs. Megha V. Gupta[1], Dr. Shubhangi Vaikole[2], Dr. PanchikattilSusheelkumar Sreedharan[3],Dr. Sachin Shinde[4], Dr. Jayant Ramesh Nandwalkar[3]

[1] Department of Computer Engineering,New Horizon Institute of Technology and Management, University of Mumbai

[2]Department of Computer Engineering, Datta Meghe College of Engineering, University of Mumbai

[3]Department of Electronics Engineering, Datta Meghe College of Engineering, University of Mumbai

[4] Department of Mechanical Engineering, Datta Meghe College of Engineering, University of Mumbai

Abstract—Nowadays it is very normal for humans to experience mild or moderate mental stress in a variety of situations. A moderate level of stress is beneficial to an individual; yet, excessive stress hurts a person's mental health and is a risk factor for suicidal ideation if ignored.Long-term stress is linked to physical health concerns, according to research.With an increasing number of individuals experiencing stress, it is critical to be able to recognize it early and assist people in addressing and resolving it before significant harm is done. Conventional methods of detecting stress levels include questioning the subject and monitoring facial expression.Stress-related questions are asked throughout the interview to have a better picture of the person's condition. When people are stressed, their brows form differently, their pupils dilate, and their blinking rate may vary.These approaches have limitations in that they may overlook stress events.Research in the stress detection domain has become quite popular. There is a scope of improvement in enhancing the accuracy of the results obtained using various methods. The use of non-invasive techniques for stress detection is quite promising. This research work proposes a system to detect human mental stress using speech signals. The human speech reflects one's mental condition.The proposed research shall analyze speech signals to recognize human mental stress using machine learningtechniques

Keywords: Stress, Speech, emotion, detection, machine learning, MFCC

1. INTRODUCTION:

Health monitoring and mental counseling systems based on artificial intelligence have received a great deal of interest in recent years because of the simplicity and efficiency of machine learning paradigms. The mental state of the people must be observed to give relevant services in various domains. We concentrate on the method for determining a user's stress level among diverse emotional states.Stress, described as the "nonspecific response of the body to any demand upon it" [1], is a fascinating emotional statefrom the standpoint of health care. This is owing to long-term stress's negative impacts, which can include everything from headaches and insomnia to an increased risk of cardiovascular disease. [2, 3, 4]. “Stress accounted for 37% of all work-related ill health cases in 2015/16”, according to the British Health and Safety Executive (HSE) [5].Every day,

people are subjected to stress.Stress reduces the effectiveness of central-peripheral regulatory systems to maintain health by making them less responsive. It has been established as a major factor in the development of chronic illnesses and the loss of productivity. It has an impact on work motivation, job performance, and life outlook. A variety of health disorders have been related to chronic stress [3]. Long-term stress exposure has been linked to risk factors such as cardiovascular disease in previous research [6, 7].Stress can create health problems either directly through physiological effects or indirectly through the development of unhealthy habits such as inadequate sleeping or eating habits, smoking, or alcohol or drug consumption [8]. Stress has also been associated with a decreased chance of living a long life [9,10].

Human mental stress is defined as an individual's unbalanced state [11] that occurs when environmental demands surpass the individual's regulation capacity [12].Stress detection is a major topic among psychologists and engineers because of its harmful implications [13]. It hasbeen utilized in lie detector tests [14], emergency call identification [15], and improving human-computer interfaces [16].People are stressed as a result of the expectations and wishes placed upon them. It gets a lot worse when they realize how difficult it will be to adjust to the circumstance [17]. Contingent upon the hour of openness to stressors, three levels of pressure might be distinguished. Acute stress is the body's natural "flight-or-fight" response in the face of adversity, and it is not dangerous.Episodicstress happens when upsetting conditions happen all the more oftentimes however disappear now and again. It is connected to arduous and rushed lives[18]. When stressors are persevering and long-standing, for example, family issues, work strain, or neediness, chronicstress happens, which is the most hurtful[19]. To stay away from stress arriving at the most elevated level and assist with lessening the dangers [20], it's critical to recognize and treat it early on, when it's still acute or episodic stress.In the last two decades, stress detection has gotten a lot of attention in a variety of domains, including medical, forensics, smart environments, teaching-learning education, human-computer interactions, emergency services, and, of course, real-time scenarios[21]. Stress is a multifaceted phenomenon that has several effects on the body and psyche.Stress has an impact on digestive functions [22], blood volume pressure [23], skin conditions [24], eating habits [25], performance [26], decision making [27], and overall health [6].Stress abbreviates telomeres — structures on the finish of chromosomes — with the goal that new cells can't develop as fast [28].Depression, stress, and anxiety disorders are anticipated to replace heart disease as the second most frequent disability in “The Global Burden of Disease” [29], a report published by the World Health Organization (WHO) in 1996.

Stress has such severe negative effects that it necessitates the use of automated detection systems. It's vital to understand that stress is primarily a physiological response to stimuli produced by the sympathetic nervous system(SNS) if you want to build a good stress detection system.Physiological changes in the body prepare the body for a physical response ('fight-or-flight').One of the most common health issues is stress; everyone has experienced it at some point in their lives, whether at work or home.Stress is the body's or mind's reaction to a physical or mental difficulty.The stress state may

affect speech distinctiveness. Every person, practically every day, feels stress, which is inextricably linked to the relationship between the environment and the individual. This stress could be a warning sign that their lives are in threat.Identification of stress is one of the most interesting study topics for psychologists and technologists alike. Stress management should start before it becomes a medical issue. Stress monitoring can help in this situation.Overall, stress detection appears to be an intriguing field of research, and a review of previous work would aid in future research. This proposed work is concerned with the recognition of human stress by using machine learning algorithms.

1.1 Machine Learning in Action

Machine learning algorithms make predictions or judgments without having to be explicitly programmed by constructing a mathematical model training data.

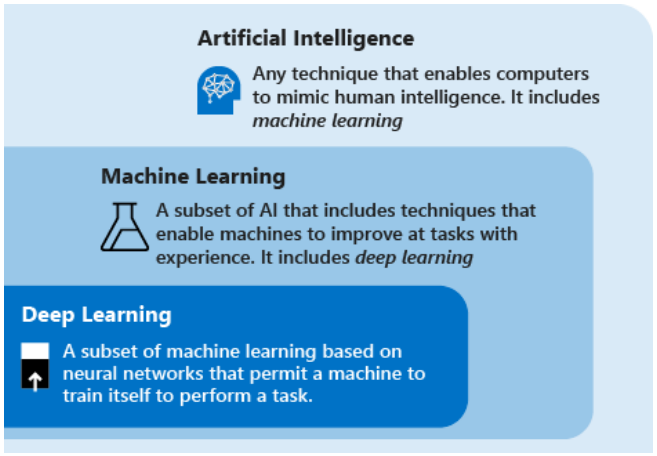

Fig. 1 Machine Learning: A subset of AI

Machine learning is used in a variety of applications, where developing a conventional algorithm for properly executing the task is troublesome or impossible, such as email filtering and computer vision. Machine learning is closely related to computational statistics, which focuses on utilizing computers to build a predictive model.Mathematical optimization studies contribute methodology, theory, and application sectors to the science of machine learning. Data mining, a related discipline of research, is concerned with unsupervised learning for exploratory data analysis.When used to tackle commercial problems, machine learning is also known as predictive analytics.Statistical and mathematical models can be used for a variety of reasons, including descriptive, predictive, and prescriptive analytics. Machine learning models are designed to extract information from data so that better business decisions may be made. Algorithmic models predict the most likely outcome for your target variable based on your training data. Data analysis is built around models. We would be stuck with elementary math (1+2=3) if we didn't have them. We wouldn't be able to discover relationships and gain insight from historical data without statistical models, and we wouldn't be able to find them without machine learning models.

1.2 CNN (Convolutional Neural Network)

It is a Deep Learning technique that can take an image as input, assign value (learnable weights and biases) to different aspects/objects in the image, and distinguish between them. A CNN requires significantly less pre-processing than conventional classification methods. While traditional techniques necessitate the hand-engineering of filters, CNNs can learn these filters/characteristics with sufficient training.The Visual Cortex's arrangement inspired CNN design, which is analogous to the Neuron Connectivity Network in the Human Brain.Individual neurons can just react to boosts in the Receptive Field, a little piece of the visual field. A gathering of such fields encloses the whole visual field.

1.3 Human-Computer Interaction:

Human-Computer Interaction(HCI) is a multidisciplinary field of study that focuses on the design of computer technology the human-machine interaction. HCI began with computers and has now grown to encompass almost every aspect of information technology design. The purpose of this research is to develop more efficient methods for humans to interface with and use computers.Computer science, sociology, and psychology are all used in HCI to improve user interfaces, improve human-human connections, and tailor computer technology to an individual's or business's needs.

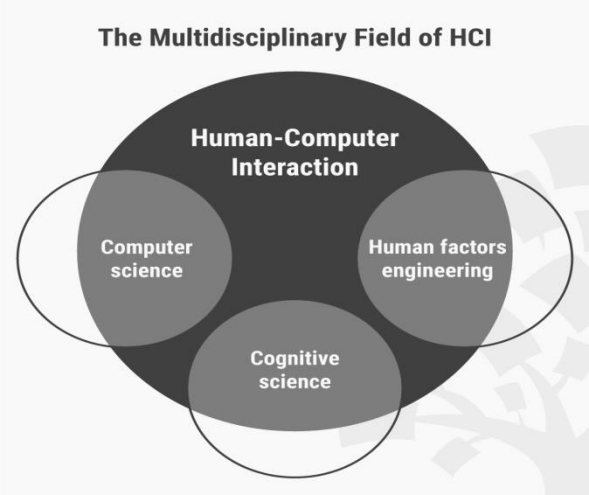

Fig. 2 The Multidisciplinary Field of HCI

2. LITERATUREREVIEW:

Classical approaches to gauging stress, such as I) interviewing the individual and asking stress-related questions to have a better understanding of their circumstances, and II) observing facial gestures. When people are stressed, they make distinct facial expressions, such as changing the shape of their brows, dilation of their pupils, or the rate at which they blink. [30] may differ - are constrained because they may overlook stress episodes.The current best quality level technique includes directing life stress interviews utilizing instruments like the Stress and Adversity Inventory, UCLA Life Stress Interview, and Life Events and Difficulties Schedule [31]. The most popular method, in turn, is to deliver brief

self-report surveys like the Perceived Stress Scale [32]. Interview-based assessments can be time-consuming and expensive, and self-report questionnaires frequently lack item specificity and validity [33].Furthermore, both approaches are retrospective and are susceptible to (sometimes unmeasured) degrees of cognitive bias and social desirability, which might have an impact on the final ratings' authenticity, reliability, and validity [32]. Existing stress detection systems are examined in-depth here, with an emphasis on how these works addressed some of the challenges.The stimuli utilized to trigger stress, the methods for measuring stress, collecting data, and using machine learning in these investigations are all outlined.To minimize long-term mental stress [34] difficulties, researchers have developed a variety of methods for detecting stress and analyzing the situations that cause it.H. Lin et al. suggested cross-media microblog data-driven automatic stress detection algorithms [35].The information is gathered by examining aspects from tweet texts and representations in the form of overlap features, such as tweeted photos, comments, and favorites.The authors were able to learn stress categories based on the attributes presented with the help of the built Neural Network.

2.1 Stress detection using Speech signal

There are several applications for stress detection using voice signals.In psychology, it is used to track the varied stress levels of patients with various stress disordersand to administer the appropriate therapies.Monitoring the stress levels of pilots, deep-sea divers, and military officers confronting law enforcement can help determine a system's safety and security.In a few criminal circumstances, stress detection is also beneficial for speaker identification, deception detection, and threat call identification [36]. A person must decide which word sequence would effectively communicate his or her desired messagewhen preparing to speak. These judgments can be influenced by stress, which can cause changes in the vocabulary, grammar, and speech tempo can be used as vocal stress indicators in the future. [37,38]. Stress, on the other hand, causes additional alterations.To propel air through the vocal folds and out of the vocal tract, the body adjusts the tension of several muscles to produce sound waves [39]. Stress raises muscular tension and breathing rate, which alters the mechanics of speech production and, as a result, the sound of speech [40,41]. Stress levels are detected from human speech using a voice-based stress detection technology called StressSense for Android phones [34]. This software was developed in a range of settings, with different speakers and scenarios.

Kevin Tomba et al. [42] detected Stress using the datasets EMO-DB, KeioESD, and RAVDESS. SVM and ANN algorithms were employed.Mean energy, mean intensity, and MFCCs were discovered to be useful features for speech analysis.

N.P. Dhole, S.N. Kale studied the RNN classification and applied it on [21] BERLIN and HUMAINE Datasets. They also applied Recurrent Neural Network on real datasets created using Audacity software. Though worked well for the audio signal, the efficiency percentage was not calculated.

Mansouri et. al. [43] designed and implemented an emotion recognition from speech signal using Wavelet and Neural Network. They used EMO-DB and SAVEE. The accuracy was discovered to be satisfactory. However, the method is time-consuming and stress detection was not considered.

2.1.1 Analysis

Overall, stress detection via speech signals appears to be an intriguing field of research, and a review of previous work would aid future research. Table 2.1 provides an overview of numerous studies reflecting the same area of interest, as well as the datasets used, the pros of the technique, and areas for improvement found.

Table 1 Analysis of previous work in the same area of interest

Title	**Dataset used**	**Technique**	**Pros of Technique**	**Scope for Improvement**
"Stress Detection Through Speech Analysis" Kevin Tomba et. al. (ICETE 2018) [42]	Berlin EmotionalDatabase (EMO-DB) the Keio University Japanese Emotional Speech Database (KeioESD) and RAVDESS	SVMs and ANNs have been chosen.	mean energy, the mean intensity, and MFCCs proved to be good features for speech analysis	Works only on audio.
"Study of Recurrent Neural Network Classification of Stress Types in Speech Identification" N.P. Dhole, S.N. Kale (IJCSE	BERLIN and HUMAINE Datasets	Recurrent Neural Network	Also works on real datasets created using Audacity software	Efficiency percentage not calculated. Works only on audio.

2018) [21]				
"Designing and Implementing of Intelligent Emotional Speech Recognition with Wavelet and Neural Network" Mansouri et. al. (IJACSA 2016) [43]	EMO-DB and SAVEE	ANN classifier	Accuracy is good	Time-consuming method. Stress detection was not considered

3. PROBLEM DEFINITION

Stress has been a major issue in society. According to global well-being research (2019) done by healthcare giant Cigna Corporation, about 82 percent of Indians feel stressed due to work, health, and financial problems.Stressful circumstances such as familial and marital difficulties, legal issues, and job loss typically precede suicidal conduct. Researchers have been attempting for decades to identify stress using various approaches in order to minimize the occurrence of suicides and depression. It is paramount to build a system to recognize stress before it becomes chronic. The proposed research work introduces a system in which, stress will be detected using Speech signals.

4. PROPOSED SYSTEMOVERVIEW:

Several stress detection techniques have been proposed for specific applications and specific requirements. However, there is a scope to improve the accuracy of the results. Although there is a lot of literature on anger, disgust, fear, happiness, sorrow, and surprise which are the primary emotions,there is not much on how to recognize and analyze them. The current work intends to contribute to the development of a human mental stress identification system against the backdrop of the aforementioned evaluation of the literature and subsequent gaps discovered from the findings of the literature review.To improve the result of stress detection using speech signals, we can use deep learning techniques

4.1 Stress detection using Speech Signal

4.1.1 Background

Emotion plays a critical role in regular interpersonal human interactions.Both rational and intelligent decision-making require this.It helps us match and grasp the emotions of others thereby allowing us to express our emotions and provide feedback to others. Emotion has a significant influence on human social interaction, according to research. Emotional expressions reveal a lot about a person's mental condition. There are advantages and cons of using speech signals to detect stress. Speech signals can be easily recognized with microphones that are not linked directly to the body, unlike bio-signal-based systems.This characteristic is important for both users and for building a massive database to be used in a stress-detection system.Speech-based stress detection systems, on the other hand, are often less accurate than biosignal-based algorithms.Despite the performance challenge that speech-based stress detection systems face, the development of neural network-based algorithms based on massive amounts of data makes such systems more promising. Human speech contains a variety of emotions that vary based on the situation. Anger, happiness, surprise, disgust, fear, sadness, neutrality, and calmness are the seven archetypal emotions that encompass all of these sentiments [44], [45]. An important issue that must be considered in a speech recognition system is the extraction of information from a voice signal to reflect a speaker's emotional state. Speech is a signal that contains temporal contextual information as well as dependencies between frames.

CNN has memory in the form of a hidden state for storing information across time and hence can easily handle contextual information like speech, that has a discrete windowed frame [46].To capture the long-term temporal peculiarities of speech, we integrate two forms of many-to-one structures with a CNN structure.The outputs are channeled to the fully-connected layers, which make their final decision with the help of a SoftMax layer.Figure 3 represents the typical Speech Emotion Detection (SED) system, which encompasses a pre-processing system, feature extraction, and classifier block.

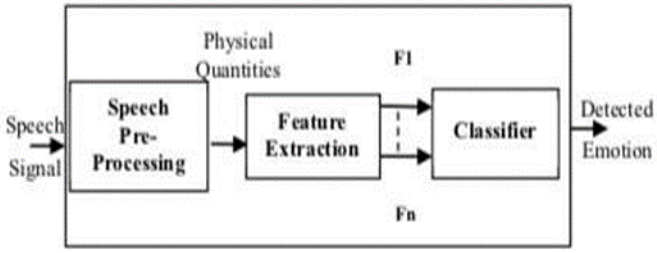

Fig. 3 Speech Emotion Detection system

After the pre-processing stage, the feature extraction block gets the raw voice/speech quantities in the speech signal.The features F1, F2,...Fn are obtained and provided to the classifier.Eventually, a person's emotions are detected using this classifier.

Speech is a signal that contains temporal contextual information as well as dependencies between frames. To enhance the efficacy and correctness of the feature extraction process,

the audio signal is first pre-processed before being sent to the feature extraction module. Filtering, Framing, and Windowing are the stages of pre-processing. After a pre-processing stage, the speech signal is used to extract raw voice quantities such as pitch, energy, and formants[47]. Filtering is a technique for decreasing noise in a speech signal caused by environmental factors or when the speech is being recorded. The objective of the pre-emphasis filter is to raise the intensity of the speech signal in higher frequencies that are attenuated during speech signal synthesis in the vocal tract.

Dataset Analysis

A. Data Collection

It's difficult to discover a database where the same speaker's voice can be classed as stressed or unstressed. We employed RAVDESS, a dataset that is multimodal and includes emotional speech and audiovisuals used in stress-related research.

B. Database Information

The RAVDESS dataset has 7356 files of a total size of 24.8 GB. This gender-balanced and validated dataset features 24 professional actors (12 females, 12 males), each vocalizing two lexically matched lines in a neutral North American accent. The different expressions used in speech are Calm, happy, sad, angry, fearful, surprise, and disgust. Each expression has two emotional intensity levels (normal and strong), as well as a neutral expression. The three modality formats accessible for all situations are Audio-Video (720p H.264, AAC 48kHz,.mp4), Audio-only (16bit, 48kHz.wav), and Video-only (no sound).The 7356 recordings were assessed ten times for emotional validity, intensity, and genuineness. 247 persons who were typical of untrained study volunteers from North America supplied ratings. The test-retest study included 72 participants in all.Emotional validity and test-retest inter-rater reliability were determined to be high. Corrected accuracy and composite "goodness" metrics are offered to help researchers in the selection of stimuli. The Speech file (Audio_Speech_Actors_01-24.zip, 215 MB) contains 1440 files 60 trials per actor x 24 actors= 1440. RAVDESS collection includes 7356 files in total (Video speech, video song, Audio Speech, and Audio song only) (2880+2024+1440+1012 files). Each of the 7356 RAVDESS files has a unique name. The filename is a seven-part numerical. The stimulus features are defined by these IDs.

Identifier	Coding description of factor levels
Modality	01 = Audio-video, 02 = Video-only, 03 = Audio-only
Channel	01 = Speech, 02 = Song
Emotion	01 = Neutral, 02 = Calm, 03 = Happy, 04 = Sad, 05 = Angry, 06 = Fearful, 07 = Disgust, 08 = Surprised
Intensity	01 = Normal, 02 = Strong
Statement	01 = "Kids are talking by the door", 02 = "Dogs are sitting by the door"
Repetition	01 = First repetition, 02 = Second repetition
Actor	01 = First actor, . . ., 24 = Twenty-fourth actor

Fig. 4Filename Identifiers

In the filename 02-01-06-01-02-01-12.mp4

● Modality is Video-only (02)

● The vocal channel is Speech (01)

● Emotion is Fearful (06)

● Emotional intensity is Normal (01)

● The statement is “Dogs are sitting by the door” (02)

● 1st Repetition (01)

● 12th Actor (12)

● Female, as the actor ID number is even.

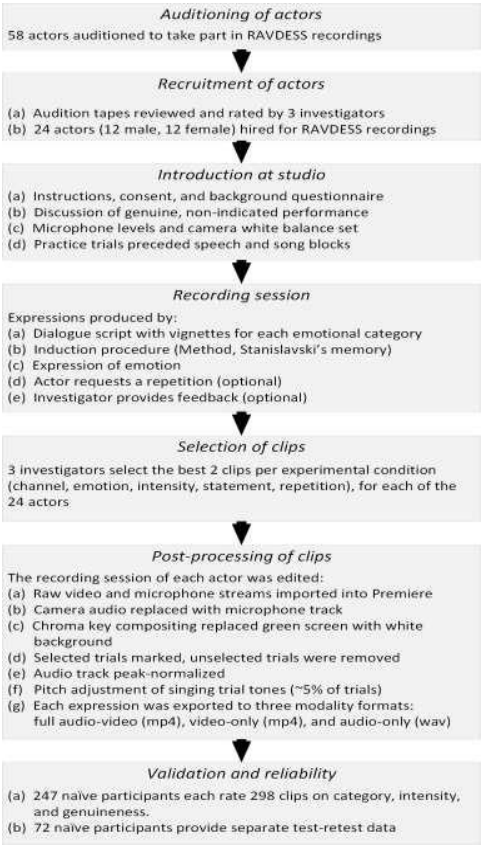

Fig. 5 Flowchart of RAVDESS Creation & Validation

The portion of the dataset of which we are going to take into consideration is the only audio part and not the video one. The main features of the dataset for analyzing the stress detection we need to only consider the sad, angry, fearful, and disgust expressions accordingly. The main sole purpose of taking only this part of the dataset files is to directly analyze the required features and then predict the stress intensity for the specific model which we have trained earlier with the neural network based on the specific approach of either the backpropagation model or the CNN network features constraints with emotional or sentimental analysis

To obtain the required result with voice amplitude and frequency manipulations available in the datasets can be implemented with better Neural Network models which can enhance the

frequency matching conversation with short memory cells and backpropagation available in the algorithm sets to extract maximum features from the sample provided for better Mel-filter bank coefficients for modulating the results as stressed or unstressed.

4.1.3 MFCC (Mel-Frequency Cepstral Coefficients)

The cepstrum can be interpreted as information on the rate of change in various spectrum bands.

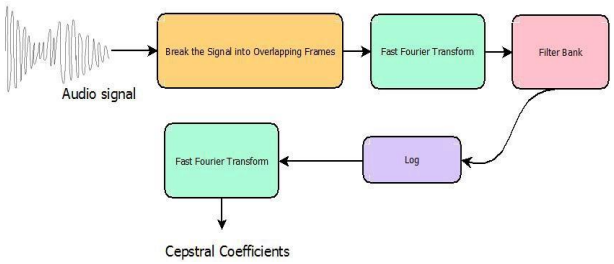

Fig. 6 MFCC Features Extraction

MFC (Mel Frequency Cepstrum) is a short-term power spectrum representation of a sound in sound processing. It is based on a nonlinear Mel frequency scale and a linear cosine transform of a log power spectrum. The coefficients that make up an MFC are called MFCCs (Mel-frequency cepstral coefficients) [44]. They're made up of the audio clip's cepstral image (a nonlinear "spectrum-of-a-spectrum").The frequency bands of MFC are evenly spaced on the Mel scale. These evenly spaced frequency bands of MFC approximates the human auditory system's response better than the cepstrum's linearly spread frequency bands. For example, in audio compression, frequency warping can help to portray sound more accurately.

4.1.4 Proposed Stress Detection System

The proposed stress detection system is a deep learning-based model that uses speech features, such as Mel-filterbank coefficients, as input.The Convolutional neural network structure uses a hard decision process to determine the user's mental stress. Here a label-based decision criterion was designed to forecast the stress status. A sad, fearful emotion represents the stressed condition, while a happy, calm emotion represents the unstressed state.A one-hot-encoding strategy was used to use these labels in the training model. Eight CNN layers and completely connected layers make up the proposed module.At each time sequence, the neural network layers collect the temporal information of the extracted features and evaluate the frame-level output f=(f1, f2,..., fT).The frame-level output is used to generate a sentence-level feature that includes all of the characteristics.These vectors are thought to offer overall information about the input features such as pitch, sample rate, formant, and energy, etc.are employed as sentence-level features fsent. This sentence-level feature is then sent to the fully connected layer.Finally, the layer outputs the average value of the output sequence favg and the layer's last frame-level output.

The block diagram for the proposed methodologyis depicted in figure 7.

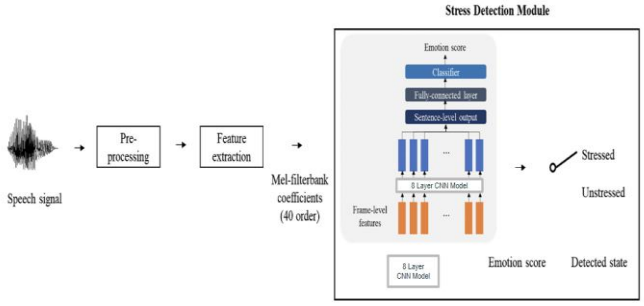

Fig. 7 Block Diagram for Stress recognition using Speech Signals.

4.1.4.1 Feature Selection

Pitch, articulation rate, energy, or MFCCs (Mel-Frequency Cepstral Coefficients) are utilized in human speech to distinguish between different emotions.There is a complete chart that shows speech quality based on six emotions (Banse and Scherer, 1996). According to this table, the five emotions of happiness, disgust, sorrow, fear/anxiety (stress), and anger can be distinguished using a combination of mean energy and mean intensity. MFCCs are used for speech analysis because they perform better in similar settings. As a result, the MFCCs, mean energy, and mean intensity are used to classify emotions.[48]

4.1.4.2 Time-Frequency Distribution Analysis

Time-frequency analysis is a signal processing technique that looks at a signal in both the time and frequency domains at the same time using various time-frequency representations. To test the validity of alternative denoising methods, we pick an appropriate threshold to separate the object signal from the laboratory background noise. The SNR (signal-to-noise ratio) is a scientific and engineering metric that compares the strength of a signal to the quantity of noise present in the environment.

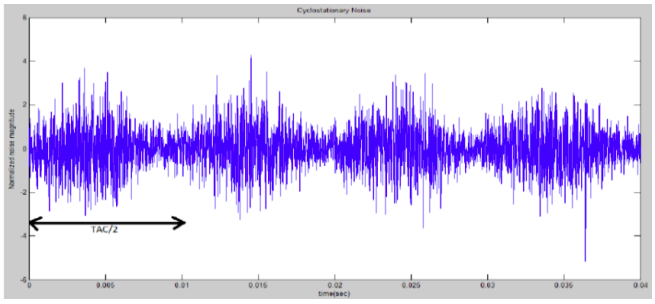

Fig. 8 Normalized noise magnitude

The signal-to-noise ratio is the ratio of the strength of a signal to the strength of background noise. Signal-to-noise ratios are expressed in decibels (dB). The logarithmic decibel scale is widely used to express signals that have a broad dynamic range.

4.1.4.3 Flowgraph

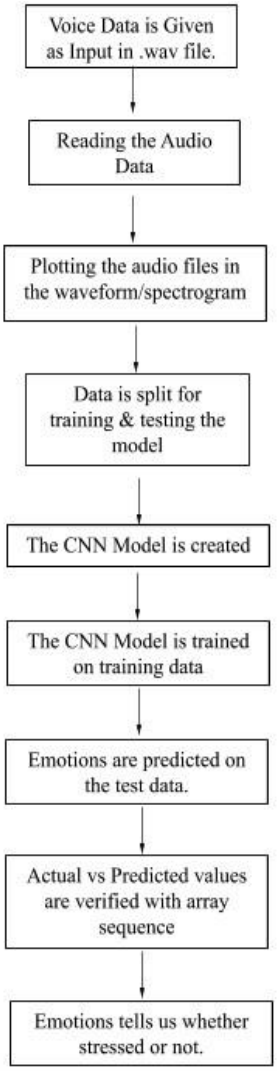

Fig. 9 Flowgraph of theSystem

5 RESULTS AND DISCUSSIONS

5.1 Screenshots of the Model Results

Speech is a signal that contains temporal contextual information as well as dependencies between frames. The screenshots are shown here below displays the Raw Audio Features and Data Generation through two methods using Noise & Pitchin figures 10 and 11.

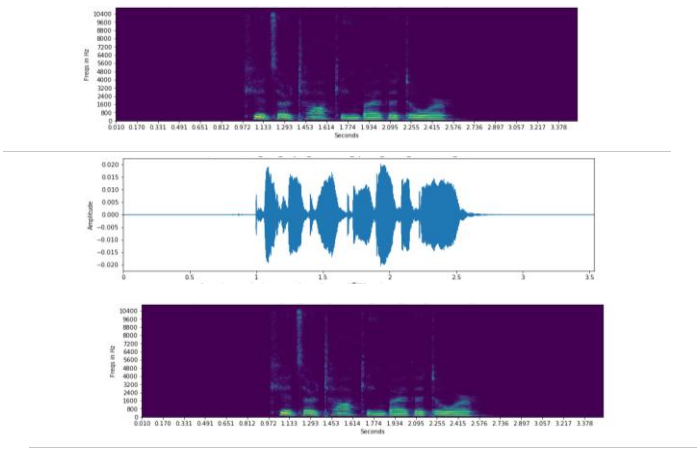

Fig. 10 Analyzing the Raw Audio features

Fig. 11 Data Generation in two-phases

The CNN layers used in this model creation for training purposes to identify the stress using the emotion labels and raw mfcc with NumPy value array in figure 12.

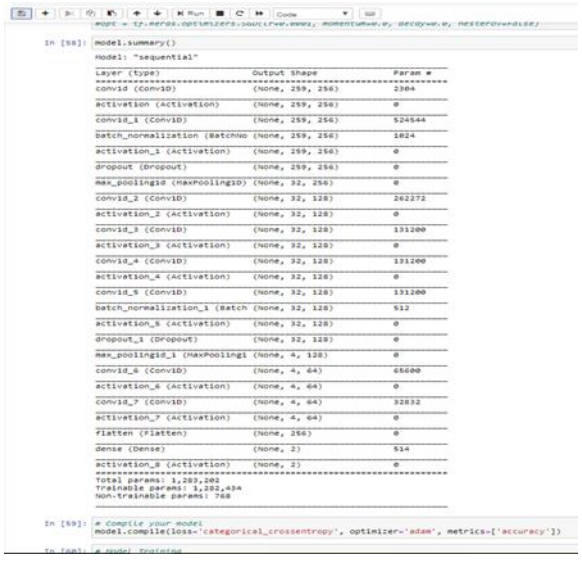

Fig. 12 Layers of CNN Model

The model training Loss and Accuracy are depicted in figure 13 and the loss of epochs are clearly illustrated in figure 14.

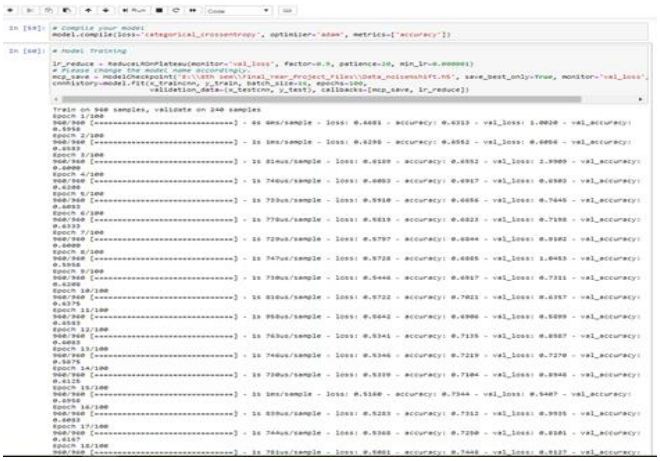

Fig. 13 Model Training Loss & Accuracy

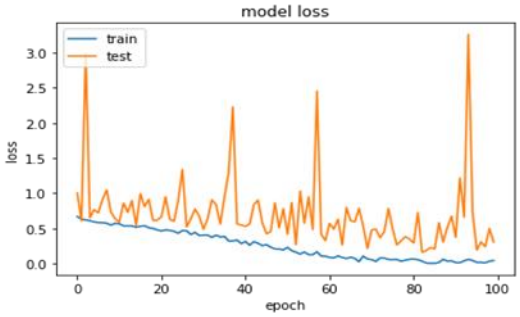

Fig. 14 Loss on epochs based on training and test data

The overall model accuracy for predicting stress is 94.25% as shown in figure 15 with some samples shown in label format i.e., Actual vs Predicted in figure 16.

```
In [88]: from sklearn.metrics import accuracy_score
         y_true = finaldf.actualvalues
         y_pred = finaldf.predictedvalues
         accuracy_score(y_true, y_pred)*100

Out[88]: 94.25

In [89]: from sklearn.metrics import f1_score
         f1_score(y_true, y_pred, average='macro') *100

Out[89]: 94.2642441006739
```

Fig. 15 Model Accuracy

Actual vs Predicted Values

```
In [76]: finaldf[10:20]
```

Out[76]:

	actualvalues	predictedvalues
10	male_positive	male_negative
11	male_positive	male_positive
12	male_positive	male_positive
13	male_positive	male_negative
14	male_positive	male_negative
15	male_positive	male_positive
16	male_negative	male_positive
17	male_negative	male_negative
18	male_negative	male_positive
19	male_negative	male_positive

Fig. 16 Actual vs Predicted Values

A confusion matrix is a table that displays how well a classification model (or "classifier") performs on a set of test data with known truth values.It enables the visualization of an algorithm's performance, as demonstrated in figure 17.

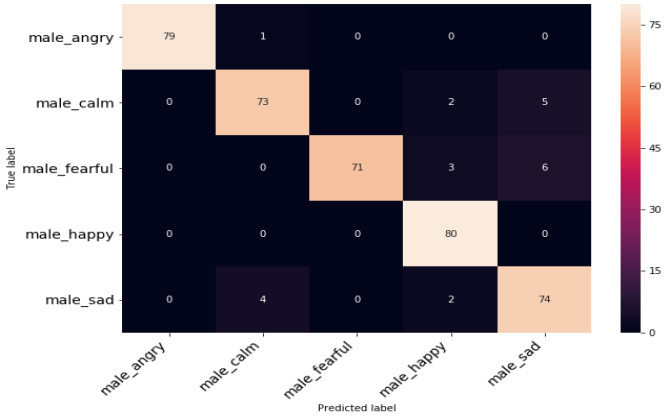

Fig. 17 Confusion Matrix

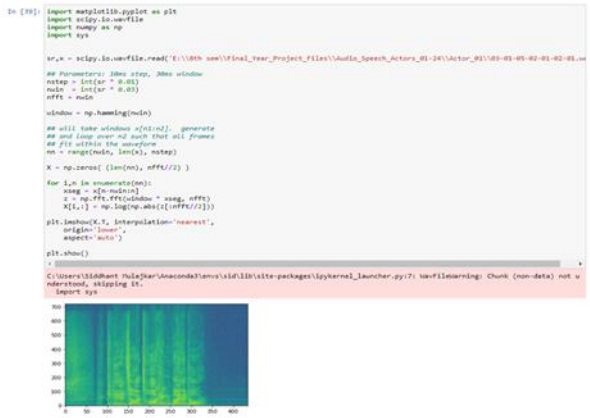

Fig. 18 Window Analysis of the audio voice

Eight CNN layers and fully connected layers make up the proposed module. After capturing the temporal information of the extracted features, the layers calculate the frame-level output at every time sequence f = (f1, f2, ..., fT). Before being sent into the fully connected layer, the frame-level output is transformed into a sentence-level feature that includes all attributes. The average value of the output sequence favg and the last frame-level output of the network layer fT are the two kinds of features that are extracted from the convolutional layer.

$$fsent = CNN(x)t=T \quad (1)$$

$$fsent = average\ (CNN(x)t=1, \ldots, T) \quad (2)$$

To retreve the output yi, the sentence-level feature is fed to the fully connected layer. Then the output yi is given as input to the classifier. SoftMax activation function and a sequential connected dense model are employed in the work. Each output can be treated as a state's probability while using the SoftMax layer. Hence, the state with the greater probability is then chosen as a final decision class, as mentioned below:

$$yi = gact\ (W(fsent\ ) + b) \qquad (3)$$

$$p\ (si\ |x) = softmax(yi) = \frac{exp\ yi}{\Sigma j exp yj} \qquad (4)$$

$$State = argmax\ si\ p\ (si\ |x) \qquad (5)$$

where s0, s1 denotes unstressed/stressed state.

For both training and testing, RAVDESS dataset containing 1440 vocal utterances of 12 male and 12 female speakers expressing emotions such as anger, anxiety, disgust, neutral, and melancholy have been considered. Table 2 shows the confusion matrices of Stressed emotion accuracy utilising Pitch/ Sample Rate, MFCC for male and female speakers, respectively.

Table 2 Stress detection System Classification Accuracy

Emotion	Pitch		MFCC(CNN)	
	Male	Female	Male	Female
Angry	59.2	62.7	98.75	99.25
Calm	22	36.4	97.5	97
Fearful	45.7	53.1	88.75	89
Happy	62.4	40	98.3	94.56
Sad	57.1	62.1	88.5	91.25
Overall Accuracy	49.28	50.86	94.26	94.3

The Stress Detection system's classification accuracy using pitch or sample rate is 52 percent for both male and female speakers, whereas the MFCC's classification accuracy is 94.33 percent for both male and female speakers, according to the findings.It is obvious from this investigation that introducing the signal raw energy operator improves the accuracy of detecting stressed emotions.

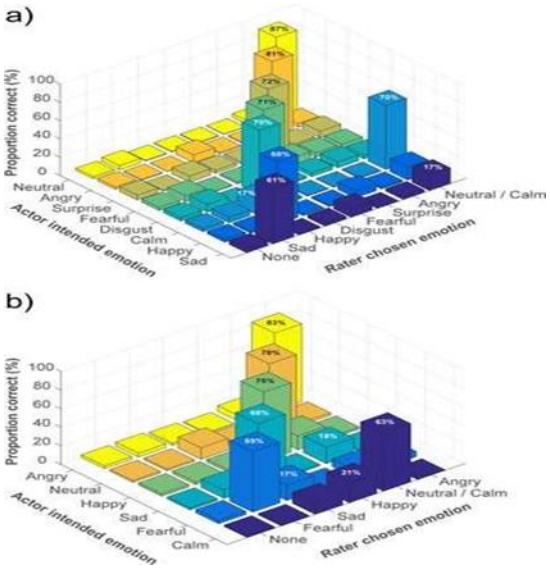

Fig. 19 Confusion Matrix of Emotional Validity

Table 3 Proposed method in comparison with baseline methods(RAVDESS dataset)

Method	Classification Accuracy using CNN
A. Christy [49]	78.3%
Mustaqeem [50]	79.5%
Prototype model	94.33%

6 CONCLUSION:

Stress is an unpleasant emotional state that leads to changes in biochemistry, physiology, and behavior.In today's world, stress is a huge issue, and workplace difficulties, such as heavy workloads and the need to adapt to constant change, are intensifying the problem. As a result of excessive stress, people are afflicted with medical issues, while businesses are losing a lot of money. As a result, monitoring stress levels is crucial in order to detect stress in its early stages and avert serious long-term consequences. The concept of stress detection arose from the need to manage chronic stress in persons. Our approach is a good starting point towards stress detection to improve the quality of life.

In this work, an audio signal is passed through a multi-step process to detect the stressed state using deep learning frameworks with a CNN structure. The stress status (i.e., stressed vs. unstressed) is recognized using a labelled classification task with emotion labels assigned.

7 REFERENCES:

[1] H. Selye, "The stress syndrome," The American Journal of Nursing, vol. 65, pp. 97-99, 1965.

[2] G. Chrousos and P. Gold. 1992. The concepts of stress and stress system disorders: an overview of physical and behavioral homeostasis. Jama 267, 9 (1992), 1244–1252.

[3] B. S. McEwen, "Central effects of stress hormones in health and disease: Understanding the protective and damaging effects of stress and stress mediators," European Journal of Pharmacology, vol. 583, pp. 174-185, 2008.

[4] R. Rosmond and P. Björntorp. 1998. Endocrine and metabolic aberrations in men with abdominal obesity about anxio-depressive infirmity. Metabolism 47, 10 (1998), 1187–1193.

[5] 2016. HSE on work-related stress. http://www.hse.gov.uk/statistics/causdis/-ffstress/index.htm. (2016). Accessed: 2017-09-06.

[6] S. Cohen, D. Janicki-Deverts, and G. E. Miller, "Psychological stress and disease," Journal of the American Medical Association, vol. 298, pp. 1685-1687, 2007.

[7] A. Steptoe and M. Kivimäki, "Stress and cardiovascular disease," Nature Reviews Cardiology, vol. 9, pp. 360-370, 2012.

[8] P. A. Thoits, “Stress and health: major findings and policy implications.,” J. Health Soc. Behav., vol. 51 Suppl, pp. S41–S53, 2010.

[9] A. Trusina, “Stress-induced telomere shortening longer life with less mutations,” BMC Syst. Biol., vol. 8, p. 27, 2014.

[10] E. S. Epel, E. H. Blackburn, J. Lin, F. S. Dhabhar, N. E. Adler, J. D. Morrow, and R. M. Cawthon, “Accelerated telomere shortening in response to life stress.,” Proc. Natl. Acad. Sci. U. S. A., vol. 101, no. 49, pp. 17312–17315, 2004.

[11] G. P. Chrousos, “Stress and disorders of the stress system,” Nat.Rev. Endocrinology, vol. 5, pp. 374–381, 2009.

[12] J. M. Koolhaas, A. Bartolomucci, B. Buwalda, S. F. D. Boer,G.Flugge, S. M. Korte, P. Meerlo, R. Murison, B. Olivier, P. Palanza, G. Richer-Levin, A. Sgoifo, T. Steimer, O. Stiedl, G. V. Dijk,M. Wohr, and E. Fuchs, “Stress Revisited: A critical evaluation of the stress concept,” Neurosci. Biobehavioral Rev., vol. 35, pp. 1291–1301, 2011.

[13] R. Lazarus, Stress and Emotion: A New Syhthesis. New York, NY, USA: Springer, 2006.

[14] I. Pavlidis and J. Levine, “Thermal image analysis for polygraph testing,” IEEE Eng. Med. Biol. Mag., vol. 21, no. 6, pp. 56–64, Nov./ Dec. 2002.

[15] L. Lefter, L. J. M. Rothkrantz, D. A. V. Leeuwen, and P. Wiggers, “Automatic stress detection in emergency (telephone) calls,” Int. J. Intell. Defence Support Syst., vol. 4, no. 2, pp. 148–168, 2011.

[16] J. Zhai and A. B. Barreto, “Realization of stress detection using psychophysiological signals for improvement of human-computer interactions,” in Proc. IEEE Southeast Con, 2005, pp. 415–420.

[17] Hunt J, Eisenberg D. Mental health problems and help seeking behavior among college students. J Adol H 2010; 46: 3-10.

[18] J. Bakker, M. Pechenizkiy, N. Sidorova, What’s your current stress level? Detection of stress patterns from GSR sensor data, in: Proceedings – IEEE International Conference on Data Mining, ICDM, IEEE, 2011, pp. 573–580. vol. 1, doi: http://dx.doi.org/10.1109/ICDMW.2011.178. <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6137431>

[19] T.W. Colligan, E.M. Higgins, Workplace stress, J. Workplace Behav. Health 21 (2) (2006) 89–97, http://dx.doi.org/10.1300/J490v21n02_07. <http://www.tandfonline.com/doi/abs/10.1300/J490v21n02_07>.

[20] N. Sharma, A. Dhall, T. Gedeon, R. Goecke, Thermal spatio-temporal data for stress recognition, EURASIP J. Image Video Process. 2014 (1) (2014) 28, http://dx.doi.org/10.1186/1687-5281-2014-28. <http://jivp.eurasipjournals.com/content/2014/1/28>.

[21] N.P. Dhole, S.N. Kale, “Study of Recurrent Neural Network Classification of Stress Types in Speech Identification” International Journal of Computer Sciences and Engineering, Paper Volume-6, Issue-4 E-ISSN: 2347-2693

[22] H. Monnikes, J. Tebbe, M. Hildebrandt, P. Arck, E. Osmanoglou, M. Rose, B. Klapp, B. Wiedenmann, and I. Heymann-Monnikes. Role of stress in functional gastrointestinal disorders. Digestive Diseases, 19(3):201{211, 2001.

[23] D. Carroll, G. D. Smith, M. J. Shipley, A. Steptoe, E. J. Brunner, and M. G. Marmot. Blood pressure reactions to acute psychological stress and future blood pressure status: a 10-year follow-up of men in the whitehall ii study. Psychosomatic Medicine, 63(5):737{743, 2001.

[24] A. Garg, M.-M. Chren, L. P. Sands, M. S. Matsui, K. D. Marenus, K. R. Feingold, and P. M. Elias. Psychological stress perturbs epidermal permeability barrier homeostasis: implications for the pathogenesis of stress-associated skin disorders. Archives of dermatology, 137(1):53{59, 2001.

[25] T. C. Adam and E. S. Epel. Stress, eating and the reward system. Physiology & behavior, 91(4):449{458, 2007.

[26] S. J. Lupien, F. Maheu, M. Tu, a. Fiocco, and T. E. Schramek. The effects of stress and stress hormones on human cognition: Implications for the field of brain and cognition. Brain and cognition, 65(3):209{37, Dec. 2007.

[27] G. Keinan. Decisionmaking under stress: Scanning of alternatives under controllable and uncontrollable threats. Journal of personality and social psychology, 52(3):639, 1987.

[28] https://www.healthline.com/health-news/mental-eight-ways-stress-harms-your-health-082713#6

[29] A. D. Lopez and C. J. Murray. The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020. Harvard School of Public Health, 1996.

[30] J A. Raij, et al., "mStress: Supporting Continuous Collection of Objective and Subjective Measures of Psychosocial Stress on Mobile Devices," In Proc. Of ACM Wireless Health 2010, San Diego, 2010. [32] H. Lu, et al., "StressSense: Detecting Stress in Unconstrained Acoustic Environments Using Smartphones," In Proc. of UbiComp 2012, Pittsburgh, 2012.

[31] Slavich, G.M. (2019). Stressnology: The primitive (and problematic) study of life stress exposure and pressing need for better measurement. Brain, Behavior, and Immunity, 75, 3–5. doi: 10.1016/j.bbi.2018.08.011

[32] Monroe, S.M. (2008). Modern approaches to conceptualizing and measuring human life stress. Annual Review of Clinical psychology, 4, 33–52. doi: 10.1146/annurev.clinpsy.4.022007.141207

[33] Shields, G.S., &Slavich, G.M. (2017). Lifetime stress exposure and health: A review of contemporary assessment methods and biological mechanisms. Social and Personality Psychology Compass, 11, e12335. doi: 10.1111/spc3.12335

[34] H. Lu, et al., "StressSense: Detecting Stress in Unconstrained Acoustic Environments Using Smartphones," In Proc. of UbiComp 2012, Pittsburgh, 2012.

[35] H. Lin et al., "Psychological stress detection from cross-media microblog data using Deep Sparse Neural Network," 2014 IEEE International Conference on Multimedia and Expo (ICME), Chengdu, 2014, pp. 1-6.

[36] Chandrasekar Vuppalapati, Mohamad S Khan, Nisha Raghu, Priyanka Veluru, Suma Khursheed “System to Detect Mental Stress Using Machine Learning And Mobile Development” Proceedings of the 2018 International Conference of Machine Learning and Cybernetics, Chengdu China IEEE 2018.

[37] G. Giannakakis, D. Grigoriadis, and M. Tsiknakis, “Detection of stress/anxiety state from eeg features during video watching,” in Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE., 2015, pp. 6034–6037.

[38] Gaikwad, Paithane “Novel Approach for Stress Recognition using EEG Signal by SVM Classifier” Proceedings of the IEEE 2017 International Conference on Computing Methodologies and Communication IEEE 2017

[39] Titze, I.R. (2000). Principles of voice production. Salt Lake City, UT: National Center for Voice and Speech.

[40] Sondhi, S., Khan, M., Vijay, R., &Salhan, A.K. (2015). Vocal indicators of emotional stress. International Journal of Computer Applications, 122, 38–43. doi:10.5120/21780-5056

[41] Zhou, G., Hansen, J.H., & Kaiser, J.F. (2001). Nonlinear feature-based classification of speech under stress. IEEE Transactions on Speech and Audio Processing, 9, 201–216. doi:10.1109/89.905995

[42] Kevin Tomba, Joel Dumoulin, Elena Mugellini, Omar Abou Khaled and Salah Hawila "Stress Detection Through Speech Analysis" Proceedings of the 15th International Joint Conference on e-Business and Telecommunications (ICETE 2018) Volume 1: DCNET, ICE-B, OPTICS, SIGMAP and WINSYS, pages 394-398.

[43] Mansouri, Mirvaziri, Sadeghi "Designing and Implementing of Intelligent Emotional Speech Recognition with Wavelet and Neural Network" International Journal of Advanced Computer Science and Applications, Vol. 7, No. 9, 2016.

[44] A. Baum. Stress, intrusive imagery, and chronic distress. Health psychology, 9(6): 653, 1990.

[45] N. Sharma and T. Gedeon. Objective measures, sensors and computational techniques for stress recognition andclassification: A survey. Computer methods and programs in biomedicine, 108(3):1287–1301, 2012.

[46] J. Lee and I. Tashev. High-level feature representation using recurrent neural networks for speech emotion recognition. 2015.

[47] I. R. Murray and Arnott J. L. Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. The Journal of the Acoustical Society of America, 93(2):1097–1108, 1993.

[48] L. Woodrow. Anxiety and speaking English as a second language. RELC Journal, 37(3):308–328, 2006.

[49] A. Christy1 · S. Vaithyasubramanian2 · A. Jesudoss1 · M. D. Anto Praveena1 "Multimodal speech emotion recognition and classification using convolutional neural network techniques" Springer Science+Business Media, LLC, part of Springer Nature 2020 29 April 2020.

[50] Mustaqeem and Soonil Kwon "A CNN-Assisted Enhanced Audio Signal Processing for Speech Emotion Recognition" MDPI Journal 28 December 2019.