International Journal of Engineering Technology and Management Sciences

Website: ijetms.in Issue: 2 Volume No.6 March – 2022 DOI:10.46647/ijetms.2022.v06i02.002

ISSN: 2581-4621

Acoustic based Stress level identification using Deep Neural architecture

S. Mahesh Reddy [1] Y. Bhanusree[2]

[1]*M.Tech,* Department of CSE, *VNR Vignana Jyothi Institute of Engineering and Technology*

[2] *Assistant professor,* Department of CSE, *VNR Vignana Jyothi Institute of Engineering and Technology*

Abstract

Stress is defined in medicine as a physical, mental, or emotional factor that generates body or mental stress. Due to stress level humans may suffer from mental, physical illness and discomfort. Unattended stress may cause serious depression which leads to instability, bipolar disorder and suicidal intentions. Stress can be identified using Electrodermal activity sensor (EDA), Respiratory sensor, Holster unit, Electroencephalogram (EEG), Electrocardiogram (ECG), Speech Identifying stress using speech is less complicated and low cost, as separate sensors are not required. The speech features like MFCC (mel-frequency cepstral coefficients), TEO (Teager energy operator), TEO-CB, TEO-PWP can be used for detection of stress. Many of the researchers in literature used Speech Under Simulated and Actual Stress (SUSAS) database for training the machine to detect stress through speech. Some of the Machine Learning(ML) algorithms like Support VectorMachine (SVM), Hidden Markov Model (HMM), K-Nearest Neighbour (KNN), Neural Network (NN) Algorithm like Multilayer perceptron (MLP) and Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), RNN-Long Short Term Memory(RNN-LSTM) are also used for stress detection in literature .In this proposed work RNN-LSTM Attention based algorithm is to be implemented to identify stress levels like High level stress, Low level stress and Neutral level stress.

Keywords: Support Vector Machine (SVM), Hidden Markov Model (HMM), K-Nearest Neighbour (KNN), Neural Network (NN)

I. Introduction

Stress has a negative impact on both mental and physical health and it is frequently a forerunner to more chronic states. Due to stress level humans may suffer from mental, physical illness and discomfort. Although stress is a natural stimulator, prolonged exposure to high levels can lead to heart attacks, hypertension etc., long-term stress has also been connected to mental health issues. Anxiety and depression are examples of health concerns [1]. Unattended stress may cause serious depression which leads to instability, bipolar disorder and suicidal intentions. Previous studies shows that there is a effect on persons voice due to stress [2]. [3] Measured the level of stress by extracting short-term variables such as pitch and energy, etc. At each sentence level, long-term variables like pitch fluctuation, speaking speed range, mean energy are considered. Stress can be identified using Electrodermal activity sensor (EDA), Respiratory sensor, Holster unit,

Website: ijetms.in Issue: 2 Volume No.6 March – 2022 DOI:10.46647/ijetms.2022.v06i02.002

ISSN: 2581-4621

Electroencephalogram (EEG), Electrocardiogram (ECG).Speech identifying stress using speech is less complicated and low cost. The speech features like MFCC (mel-frequency cepstral coefficients), TEO (Teager energy operator), TEO-CB, TEO-PWP can be used for detection of stress.

Because of convenience, effectiveness of machine learning (ML) algorithms demands more in artificial intelligence. It aids especially in healthcare monitoring and also psychological treatment systems has grown in recent years. The user's mental state must be observable in order to deliver relevant services in these areas. We focus on an approach for detecting the user's stress status using solely voice patterns, among other emotional states [4]. The usage of speech signals to identify stress has both pros and cons. Existing approaches to evaluate stress with This stress-related change in the quality and pattern of speech acoustics is used to quantify the level of stress a person is currently feeling. This can be accomplished by evaluating numerous characteristics, including the fundamental frequency, as shown in Figure 1. Machine learning algorithms are used to analyze these variables and provide a real-time indicator of a person's stress state. The associated continual stress signal can be used to determine a person's ongoing health concern, which interview-based or self-report techniques cannot achieve [5-7].[8] suggested an approach for classifying low-level data such as mel frequency cepstrum coefficients (MFCC) and voiced pitch using the support vector machines (SVM) algorithm.The purpose of the project research is to effectively assess an individual's stress level using physiological data collected during stressful scenarios.

This type of detection can aid in stress monitoring and the prevention of harmful stress-related disorders.

For stress detection and recognising a person as worried or unstressed (also normal, amused, or stressed), many machine learning and deep learning algorithms are applied.

Understanding the structure and format of the publicly available dataset, cleaning and transforming data into a set eligible for machine learning and deep learning classification methods, exploring and constructing various classification models, and comparing them are all steps taken to achieve this goal.

Saskia Koldijk et al. [9] developed automatic classifiers to examine the relationship between working conditions and mental stress-related conditions from sensor data: body postures, facial expression, computer logging, and physiology (ECG and skin conductance). They discovered that when similar users were subgrouped and models were trained on specific subgroups, the performance of the specialised model was equal to or better than that of a generic model in almost all cases. Among the most useful modalities for distinguishing between stressor and non-stressor working conditions, posture provides the most critical information. Adding data about one's facial expressions could improve performance even more. Using an SVM classifier, they achieved an accuracy of 90%. [10] the work focuses on detecting each individuals stress levels by using ML and DL techniques and

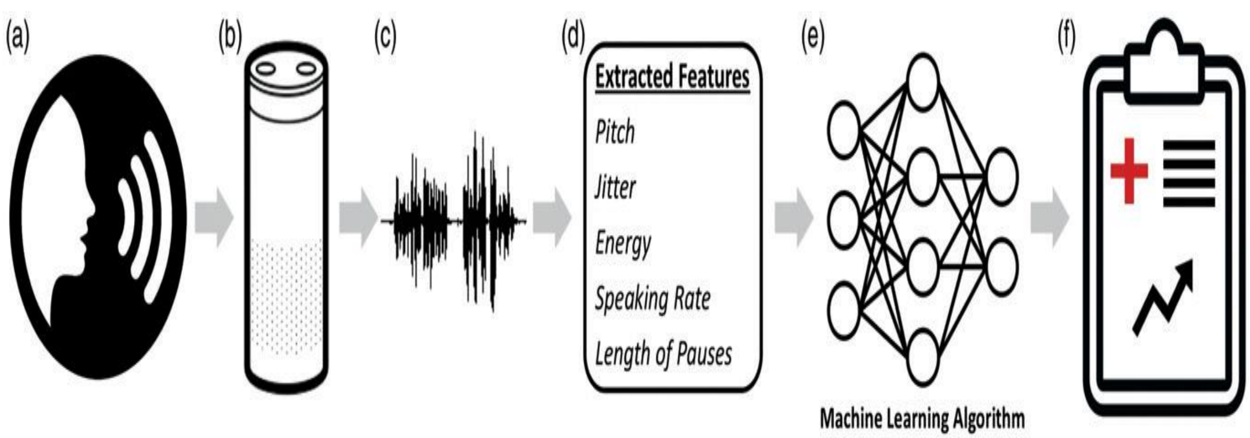
Website: ijetms.in Issue: 2 Volume No.6 March – 2022 DOI:10.46647/ijetms.2022.v06i02.002

ISSN: 2581-4621

for this the authors used multimodal dataset and achieved an accuracy of 95.21%. they discovered that sensor data can better predict the subjective variable'mental effort' than, say, 'felt stress.' A study of multiple regression approaches revealed that a decision tree is the best predictor of mental effort (correlation of 0.82). The most useful information comes from facial expressions, followed by posture. Individual variances are important to consider, especially when measuring mental states. When we build models on specific subgroups of comparable users, a specialised model performs as well as or better than a generic model[11]. This study presents an integrated Physiological Sensor Suite (PSS) based on QUASAR's novel non-invasive bioelectric sensor technologies, which will enable a completely integrated, noninvasive physiological sensing technique for the first time. The PSS is a cutting-edge multimodal array of sensors that, when combined with an ultra-low-power personal area wireless network, provide a comprehensive body-worn system for real-time monitoring of subject physiology and cognitive condition[12].

In this work, we aimed to develop a long short-term memory-recurrent neural network (LSTM-RNN) that learns and also interprets speech signals. It uses a feature that stores the overall spectral information of the signals in order to detect stress. We develop a neural network that learns and interprets speech signals. It uses a feature that stores the overall spectral information of the signals in order to detect stress [13]. An LSTM structure can store information about a long-term state in a hidden state. It can also handle certain details of speech, such as its frequency and duration [14]. For the study, we collected a multi-modal database of speech, video, and bio-signals from 56 subjects. We were able to obtain signals in both stress and non-stressful conditions [15].

The following is a description of the paper's structure. Section 2 provides background information on the study as well as specifics on the data collection sets. In Section 3, details of the models, utilized for experiments, and in Section 4, gives our findings and recommendations for further research.



International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.6 March – 2022 DOI:10.46647/ijetms.2022.v06i02.002
ISSN: 2581-4621

Figure 1. Assessing stress using speech. (a) Speaker (b) device for recording (c) captured audio signal. (d) Features of obtained audio signal which are inputs to (e(ML algorithm)) (f) Clinical chart to assess patients' potential disease risk.

Table 1: Reviewed articles along with methodology &performance measures.

s.no	**Title**	**Author**	**Methodology & Performance**
1	Stress measurement using speech: Recent advancements, validation issues, and ethical and privacy considerations	George M. Slavich , Sara Taylor and Rosalind W. Picard	smart phones and smart speakers have been used assess stress.In this open source program like open smile is used. Mel frequency cepstral coefficients is used for stress speech feature extraction.Machine Learning algorithms are used.
2	Stress detection from speech signalUsing mfcc, svm and machine learningTechniques	Dr. BageshreePathak ,Chinmayi Dhole ,HarshadaHajare ,MrunalZambare	In this paper we have seen a system that can recognize whether an individual is in stress or non-stress, given audiowith various techniques like speech signal processing, machine learning, human psychology.SVM algorithm is used for classification.Mel frequency cepstral coefficients is used for speech feature extraction.The database used in this are CPR departments trainee officers and media coverage and Youtube videos.
3	A Deep Learning-based Stress Detection Algorithmwith Speech Signal	Hyewon Han,Kyunggeun Byun,Hong-Goo Kang	This paper proposes a deep learning-based psychological stress detection algorithm using speech signals. Multimodal database is used for Feature extraction.SVM algorithm is used. LSTM-RNN layers and fully connected layers.The algorithm used are Long Short Term Memory(LSTM) and feed forward networks.Using the proposed algorithm we achieved 66.4% accuracy.The database used in this are Multi modal data base.In the feature extraction mel - filterbank coefficients were used.
4	Stress Speech Identification Using Various NeuralNetworks	Mrs. N.P. Dhole , Dr.S.N. Kale	The database used in this work are Berlin database and Humaine database as benchmark datasets.Mel frequency cepstral coefficients is used for stress speech feature extraction.The algorithms used in this are Support Vector machine(SVM),Radial Basis Functions(RBF),Recurrent Neural Networks(RNN)and Multilayer Perceptron (MLP), among all these MLP is the best identifier for real datasets for stressin speech signals.In this Audacity software is developed.
5	Research on Speech Under Stress Based on	Xiao yao ,ning xuXiaofengliuaiminJiang ,	This paper proposes a method for a research of speech under stress based on a physical model and

Website: ijetms.in Issue: 2 Volume No.6 March – 2022 DOI:10.46647/ijetms.2022.v06i02.002

ISSN: 2581-4621

	GlottalSource Using a Physical Speech Production Model	and xuewuZhang	glottal flow.
6	Attention-Based LSTM for psychological stressDetection from spoken language using distantsupervision	Genta Indra Winata,OnnpPepijnKampman,PascaleFung	This paper proposes a long short-term memory(LSTM) with attention mechanism to classify psychological stressfrom self conducted interview transcriptions.The bidirectional LSTM model with attention is found to be the best model in terms of accuracy and f-score.The major work done by this paper is that unlabeled data collected from twitter can improve the classificationperformance on our interview transcriptions corpus and that applying an attention mechanism helps the model toeffectively choose important words.
7	A Machine Learning Approach for StressDetection using a Wireless Physical ActivityTracker	B. Padmaja, V. V.Rama Prasad and K. V.N. Sunitha	It provides an effective method for the detection of cognitive stress levels using data provided from a physical activitytracker device developed by FITBIT.This system was used and evaluated in a real-time environment by taking data from adults working in IT and othersectors in India.We are currently working on studying the stress levels (low, medium, high) among professionals using the datacollected from a wireless physical activity tracker developed by FITBIT.This paper aims to detect the stress levels of an individual.The Akaike information criterion (AIC) is used to quantify the relative quality of logistic models for a given data set.
8	Detecting stress and Depression in adults with Aphasiathrough speech analysis	Stephanie Gillespie,ElliotMoore,JacquelineLaures-gore,MatthewFarina,ScottRussell,Yash-Yee Logan	This paper proposes to use speech analysis as an objective measure of stress and depression in patients with aphasia.The algorithms used are support vector machines(SVM) and linear support vector regression model(linear-SVR).Prosodic , spectral TEO and glottal features were extracted from voiced sections of speech.Among all these features Teager energy operator- amplitude modulation performed the best in predicting stress.Aphasia database is used in this work.
9	Stress Detection with Machine Learning and DeepLearning using Multimodal Physiological DataPramod	Bobade, VaniM.	In this paper different Machine Learning and Deep Learning techniques are used for stress detection.Multimodal dataset is used.Stress states are taken from WESAD dataset.The accuracy achieved is up to 95.21% .The main aim of this work is to automatically detect the stress conditions of an individual.
10	Stress Detection through Speech Analysis usingMachine Learning	Dr. S. Vaikole, S.MuJayaswal, S.Dhas.lajkar, A. More,P.	In this paper we studied a deep learning-based psychological stress detection model using speech signals.The proposed module is composed of eight CNN layers and fully connected layers.The database used in this paper are Ryerson Audio-Visual Database of Emotional Speech andSong(RAVDESS).In the feature extraction mel -filterbank coefficients were used.By using MFCC

			Accuracy obtained is 94.33%.

II. Methodology

Our goal is to evaluate whether or not someone is stressed based on a speech. We propose a trainable embedding layer for the LSTM model, their vectors must ultimately create both stress and de-stress clusters. The temporal potent of phrases can be captured using LSTMs can be seen in fig 2. A (RNN) is a DL network along with recurrent structure. As a result, time-series data, speech signals, can be represented effectively. Input, output vectors, with hidden state ht and this state from previously step ht1 make up the RNN network. One recurrent layer in the LSTM network procreate the embedding vector bt for one word at one time t to find hidden state ht.

$$\overrightarrow{h_t} = LSTM(b_t), t \in [1, T]$$

A next layer receives all hidden states.Due to the fact that not all words lend similarly to the stress category, we introduced this layer. $u_t = tanh(W\overrightarrow{h_t} + b)$

is used to calculate the word significance vector ut. A softmax function

$$\alpha_t = \frac{exp(u_t^T u)}{\sum_t exp(u_t^T u)}$$

calculates the normalised word weight t. The weight sum of hidden states with t is corresponding weights is the total of all the information in the phrase v[16].

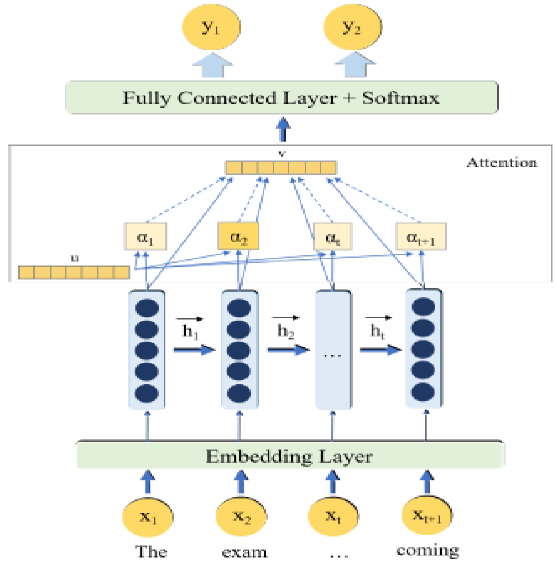


Fig. 2. Attention-based LSTM architecture

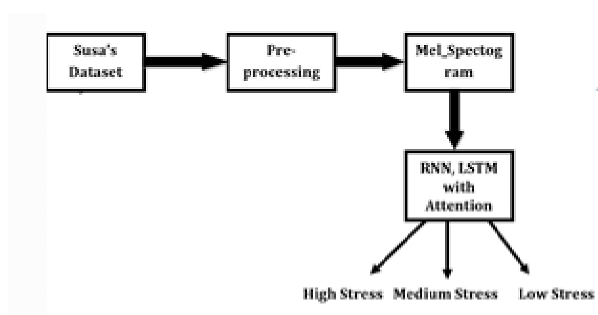


Fig 3. Proposed block diagram

III. Proposed algorithm

Figure 3 shows proposed block diagram. The voice features, extracted from feature extraction module and fed to a deep-learning classifier. RNN-LSTM model consists of completely connected layers. The LSTM layers takes temporary information from extracted features and time sequence for

International Journal of Engineering Technology and Management Sciences

Website: ijetms.in Issue: 2 Volume No.6 March – 2022 DOI:10.46647/ijetms.2022.v06i02.002

ISSN: 2581-4621

frame level output f=(f1, f2,..,fT) is calculated. This converts to sentence-level features, which enhances their characteristics and fed into fully connected layer. The LSTM has 2 features namely *favg*for output sequence value andfT for frame-level output. *Fsent*are sentence-level features which has entire information.

$$f_{sent.} = LSTM(x)_{t=T} \tag{1}$$

$$f_{sent.} = average(LSTM(x)_{t=1,\dots,T}) \tag{2}$$

The output yi is obtained when sentence-level features are fed into fully connected network. As a classifier, we used a function called softmax and SVM in our work. When the softmax layer is used, each output could be viewed as a likelihood for every state. As a result, the highest probability condition is chosen to be final decision class[11].

$$y_i = g_{act.}(W(f_{sent.}) + b) \tag{3}$$

$$p(s_i|x) = \text{softmax}(y_i) = \frac{\exp y_i}{\sum_j \exp y_j} \tag{4}$$

$$State = \underset{s_i}{\text{argmax}}\ p(s_i|x) \tag{5}$$

Here s0, s1 indicates unstressed/stressed condition.

From fig 4,The proposed system was done using susa’s database. The softmax classifier was used to train the linear SVM, and the cross-entropy loss function was used as a training criterion to optimize the log-likelihood log(p(s |x)).For updating the model, an Adam optimizer with β1 = 0.9, β2 = 0.99, ϵ = 10−8 are used. Four stress-detection modules were trained and compared in this experiment, each using distinct sentence-level characteristics and classifiers. The performance of every classifier on the testing set is shown in Table 1. the LSTM model, sentence-level feature outperformed the other three models. Pre-processing of speech is a more crucial stage in the creation of an automatic speech recognition system in signal processing.

Signal Pre-processing ,employs a noise removal algorithm, which results in the plot of the original signal. The Librosa library is imported from Python for this purpose. Figure 5 depicts the original voice signal, with time in milliseconds on X-axis and amplitude values on Y-axis. Figure 6 depicts that preprocessed signal from Fig 5 at the default sampling rate of 22.05 kHz. This converts the audio channel from stereo to mono[17]. During feature extraction, raw data is first reduced to a more manageable size for processing.

So we employed the Mel-frequency Cepstral Coefficients (MFCC) feature extraction technique for that. Figure-7 depicts a plot of MFCC features, X-axis shows with MFCC coefficients and Y-axis shows frame index, indicating the number of frames. For our dataset, we used 13 MFCC coefficients per frame.

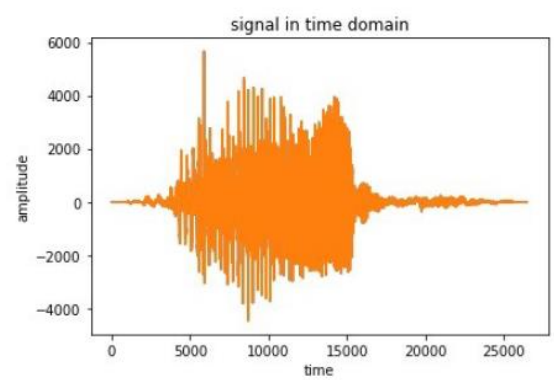


Fig. 4: Module for Stress Detection Design

International Journal of Engineering Technology and Management Sciences

Website: ijetms.in Issue: 2 Volume No.6 March – 2022 DOI:10.46647/ijetms.2022.v06i02.002

ISSN: 2581-4621

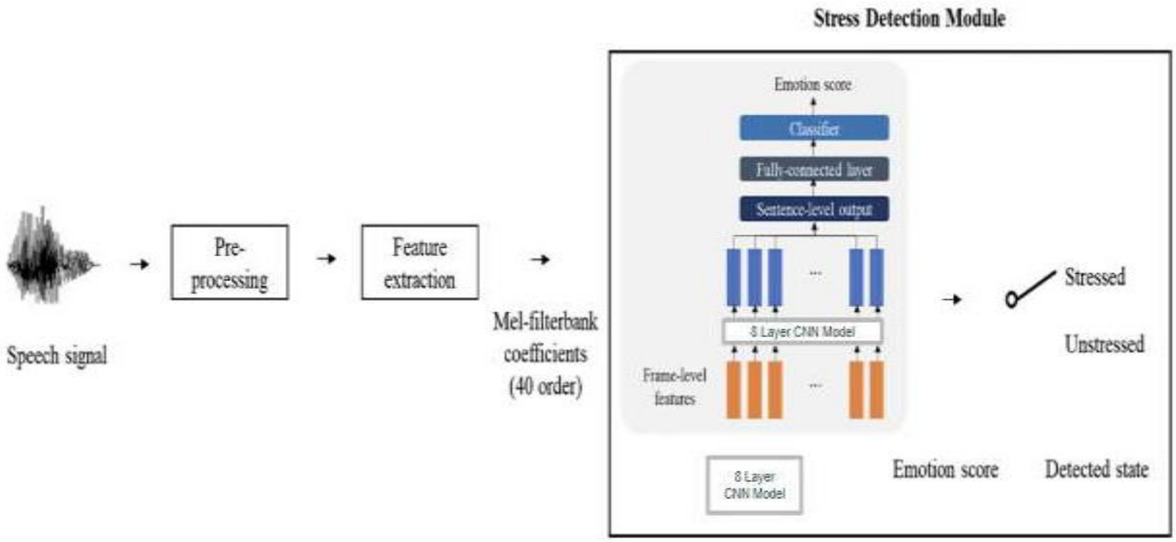


Fig 5: original signal

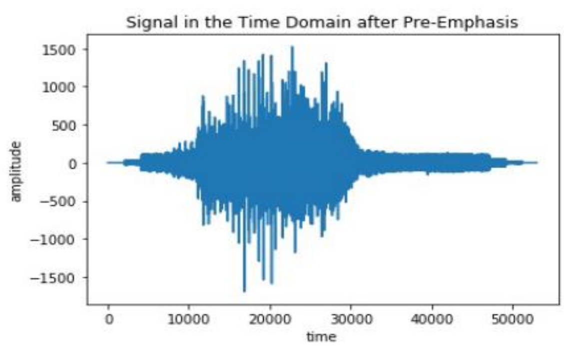


Fig 6: pre-processed signal

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.6 March – 2022 DOI:10.46647/ijetms.2022.v06i02.002
ISSN: 2581-4621

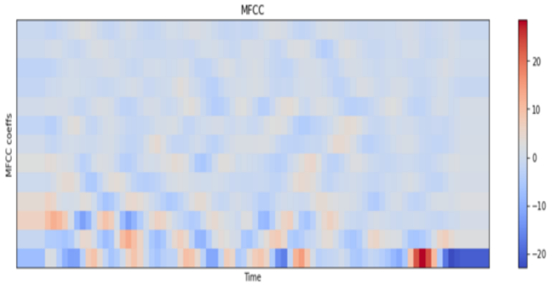


Fig 7: MFCC feature extraction

IV. Results

The proposed study has identified two classification tasks for stress detection based on a person's emotional states. The task was divided into three categories: amusement, baseline, and stress. Second, the amusement and baseline states were combined to create a non-stress class, and a binary classification task: stress vs. non-stress was established.

Figure 8 depicts the loss model of RNN-LSTM and SVM classifier of proposed model. On X-axis we plotted epoch of signal, on Y-axis we defined loss of a model. To train the proposed model we used 90 recorded samples and achieved 91.25% accuracy by using SVM classifier.[18]

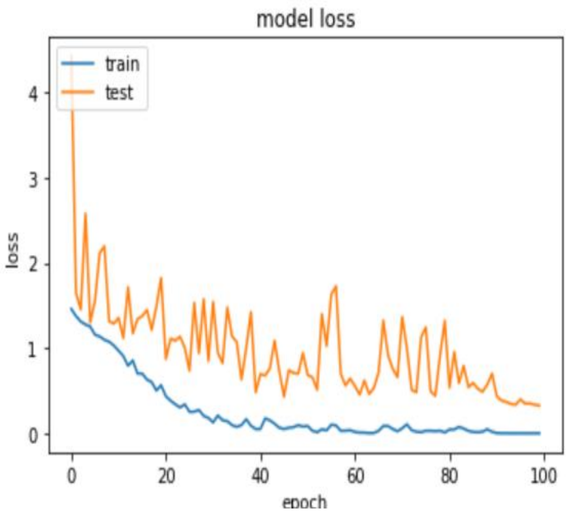


Fig 8: Loss model

Confusion matrix is obtained from classifier performance using validation data of svm classifier and given below:

Table 1: confusion matrix for stress detection module

Total samples: 90	Stress	Non-stress
Stress	48	5
Non-stress	4	33

The classification accuracy of the Stress Detection system utilising pitch or sample rate is 62 percent for both male and female speakers, and MFCC is 96 percent for both male and female speakers, according to the data. It is obvious from this investigation that adding the signal raw energy operator improves the accuracy of detecting strained emotions.[19]

V. Conclusion

In this project, we created an algorithm to determine whether or not a person is stressed. To implement Python code, we used Python 3.7 software and the Spyder IDE compiler. Stress is an adverse emotional situation that causes physiological changes. In three steps, we collected voice data and audio data and built stress-detection model using deep learning with LSTM structures to detect stressed state using only voice signals. We will also include the model described now into a virtual therapist platform [20], which include ASR output. The system is alerted to the user's stress, which is what it treats with stress management recommendations and exercises.

As a result, the approach can most likely be developed to a superior conceptual multi-modal based strategy to improve detection accuracy even more. Professional measurements of the variance in cortisol levels in each raw audio stage could yield more trustworthy experimental results. In the future, we will take into account all of these factors in order to develop a more accurate stress detection model.

References

1. G. I. Winata, O. P. Kampman and P. Fung, "Attention-Based LSTM for Psychological Stress Detection from Spoken Language Using Distant Supervision," *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 6204-6208, doi: 10.1109/ICASSP.2018.8461990.
2. N. Sharma and T. Gedeon. Objective measures, sensors and computational techniques for stress recognition and classification: A survey. Computer methods and programs in biomedicine, 108(3):1287–1301, 2012.
3. Potluri T., Jahnavi S., Motupalli R. (2021) Mobilenet V2-FCD: Fake Currency Note Detection. In: Luhach A.K., Jat D.S., Bin Ghazali K.H., Gao XZ., Lingras P. (eds) Advanced Informatics for Computing Research. ICAICR 2020. Communications in Computer and Information Science, vol 1393. Springer, Singapore. https://doi.org/10.1007/978-981-16-3660-8_26 .
4. Han, Hyewon& Byun, Kyunggeun& Kang, Hong-Goo. (2018). A Deep Learning-based Stress Detection Algorithm with Speech Signal. 11-15. 10.1145/3264869.3264875.
5. Suresh C., Ravikanth M., Srivani B., Satish T. (2021) Cognitive IoT-Based Smart Fitness Diagnosis and Recommendation System Using a Three-Dimensional CNN with Hierarchical Particle Swarm Optimization. In: Gupta D., Hugo C. de Albuquerque V., Khanna A., Mehta P.L. (eds) Smart Sensors for Industrial Internet of Things. Internet of Things (Technology, Communications and Computing). Springer, Cham. https://doi.org/10.1007/978-3-030-52624-5_10.
6. Hansen, J.H.L., & Patil, S.A. (2007). Speech under stress: Analysis, modeling and recognition. In Muller C. (Ed.), Speaker classification I. Lecture notes in computer science, vol. 4343 (pp. 108–137). Berlin, Heidelberg: Springer.
7. Ravi Kanth Motupalli, Dr. O. Naga Raju “Modelling Disaggregated Smart Home Bigdata for Behaviour Analytics of a Human using Distributed architectures” in journal of critical reviews, Volume 7, Issue 19, 2020, ISSN: 2394-5125, Page no: 522-530
8. Vignolo L. Schlotthauer G. Colominas M.A. Rufiner H.L. Sharma, R. and S.R.M. Prasanna. Empirical mode decomposition for adaptive am-fm analysis of speech:a review. Speech Communication, 88:39–64, 2017.
9. Ravi Kanth Motupalli, Kota Chaitanya Kumar “Arduino Uno based Smart Helmet with Protective Bike System Using IoT” in Journal of Advanced Research in Dynamical & Control Systems, Vol. 12, No. 9, 2020, ISSN 1943-023X, Page no: 27-33.

Website: ijetms.in Issue: 2 Volume No.6 March – 2022 DOI:10.46647/ijetms.2022.v06i02.002

ISSN: 2581-4621

10. K. Manley. Comparative study of foreign language anxiety in korean and Chinese students. 2015.
11. Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alexander J Smola, and Eduard H Hovy, “Hierarchical attention networks for document classification.,” in HLT-NAACL, 2016, pp. 1480–1489.
12. Ravi Kanth Motupalli, Dr. Joel Sunny Deol Gosu, Dr. PullaguraPriyadarsini, “A Hybrid Approach for the Analysis of Feature Selection using Information Gain and BAT Techniques on The Anomaly Detection” in Turkish Journal of Computer and Mathematics Education, Vol.12 No. 5 (5th April ,2021), 656-666.
13. Mara, Pooja, and Ravi kanthMotupalli. "Blockchain-based model to track and verify official certificates."
14. Genta Indra Winata, OnnoKampman, Yang Yang, Anik Dey, and Pascale Fung, “Nora the empathetic psychologist,” Proc. Interspeech 2017, pp. 3437–3438, 2017.
15. Ravi Kanth Motupalli, Dr. O. Naga Raju” Integration of SQL Modelling and Graph Representations to Disaggregated Human Activity Data for Effective Knowledge Extraction” inPSYCHOLOGY AND EDUCATION (2020) 57(8): 975-984, ISSN: 00333077.
16. Ravi Kanth Motupalli, Dr. O. Naga Raju “A Systematic Review on Modelling of Human Activity Patterns in Smart Homes” in Science, Technology and Development, Volume VIII Issue XII DECEMBER 2019, ISSN: 0950-0707, Page no: 942-950.
17. Matthews, Robert & McDonald, Neil & Hervieux, Paul & Turner, Peter &Steindorf, Martin. (2007). A Wearable Physiological Sensor Suite for Unobtrusive Monitoring of Physiological and Cognitive State. Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference. 2007. 5276-81. 10.1109/IEMBS.2007.4353532.
18. Virginia Sandulescu, Sally Andrews and David Ellis, Stress Detection using wearable physiological sensors, Springer International Publishing Switzerland, 2015.
19. Md Fahim Rizwan, Rayed Farhad and Farhan Mashuk, "Design of a biosig-nal based Stress Detection System using machine learning techniques", *2019 International Conference on Robotics Electrical and Signal Processing Techniques (ICREST)*.
20. Saskia Koldijk, Maya Sappelli and Suzan Verberne, "The SWELL Knowledge Work dataset for stess and user modeling research", *ICMI*, 2014.
21. Mohamed A.R.Graves,A.andG.Hinton. Speech recognition with deep recurrent neural networks. In Acoustics, speech and signal processing (icassp), 2013 IEEE International Conference on, pages 6645–6649. IEEE, 2013.