ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

DEVELOPING MACHINE LEARNING MODEL FOR FINDING STRESS LEVELS USING SPEECH PATTERNS

Billa Pravalika

21N81A6774

Computer Science and Engineering (Data-Science)

Sphoorthy Engineering College,

Nadergul, Hyderabad,501510

billapravalika21@gmail.com

C. Harika

21N81A6787

Computer Science and Engineering (Data-Science)

Sphoorthy Engineering College,

Nadergul, Hyderabad,501510

harikachittanoori@gmail.com

Abdul Aziz

21N81A67A3

Computer Science and Engineering (Data-Science)

Sphoorthy Engineering College,

Nadergul, Hyderabad,501510

Shahaziz7864@gmail.com

C. Nishanth

21N81A67B8

Computer Science and Engineering (Data-Science)

Sphoorthy Engineering College,

Nadergul, Hyderabad,501510

chegurinishanth@gmail.com

Mrs Y Priya

Assistant Professor

Computer Science and Engineering (Data-Science)

Sphoorthy Engineering College,

Nadergul, Hyderabad,501510

r.priya@sphoorthyengg.ac.in

ABSTRACT: An interactive web application was developed to analyse audio signals in real time. It captures live input from the microphone and processes it efficiently using numerical operations. The system extracts important characteristics from the audio, enabling further analysis such as classification or pattern recognition. To ensure smooth performance, audio data is handled asynchronous in the background, avoiding interruptions in the user experience. Timing mechanisms are used to synchronize tasks and monitor

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

processing speed. Captured audio is stored in a standard format for playback or further study. The application also leverages previously trained models aware systems across various domain.

KEY WORDS

Real-time audio pre-processing , Interactive web application, Audio signal analysis, Feature extraction, data handling, Timing and synchronization, Audio storage, Machine Learning, Asynchronous

1.INTRODUCTION

In today's fast-paced world, stress significantly impacts mental and physical well-being. Traditional stress detection methods, such as self-reporting or physiological sensors, have limitations. This project develops a machine learning model to detect stress levels from speech patterns in real-time, analysing tone, pitch, tempo, and pause patterns. By using feature extraction techniques like Mel-frequency cepstral coefficients (MFCCs) and classification algorithms like Random Forest or deep learning networks, the system identifies subtle cues linked to stress. The project includes a user-friendly interface, making it accessible to non-technical users. With potential integration into mobile applications or smart home devices, users can receive instant feedback about their stress levels, enabling proactive steps to interpret audio features and generate predictions. All components work together to deliver a responsive and intelligent interface. This design allows users to interact with and understand sound in a dynamic and insightful way. It offers a foundation for building intelligent, sound toward relaxation or The model is designed with privacy and security in mind, using secure data handling and anonymization protocols. This innovative approach contributes to affective computing and opens pathways for early stress detection and mental health support, empowering individuals and mental health professionals with realtime insights. The system's adaptability and continuous learning capabilities ensure its effectiveness across diverse populations and environments. It can be fine-tuned for different demographics, languages, and speaking styles. This flexibility makes it a valuable tool for various applications. The project's impact extends beyond technical innovation, contributing to societal wellbeing. By providing real-time stress insights, it enables individuals to take proactive measures. This can lead to improved mental health outcomes and overall quality of life.

2.LITERATURE REVIEW

[1] Speaker Embeddings as Individuality Proxy for Voice Stress Detection-The study focuses on detecting stress in voice using machine learning.

It builds on existing models but aims to improve generalization across datasets and stress types. Traditional models use long audio (10–30 seconds); this work reduces input to 3–5 seconds. The core model used is Hybrid BYOL-S/CvT, a self-supervised audio representation model. These embeddings are concatenated with audio features to personalize predictions. They use Support Vector Machines (SVMs) and Multi-Layer Perceptron's (MLPs) for classification. Testing is done across 5 datasets, 111 speakers, 9 languages, with cognitive and physical stress. Audio is chunked into 3-5s segments to simulate real-time use cases. ECAPA speaker embeddings improve model performance significantly more than Resemble. Combining audio and speaker embeddings gives the best results. ECAPA also

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

captures useful paralinguistic cues beyond speaker identity. Shorter chunks reduce performance, but speaker embeddings help maintain accuracy Framewise embeddings did not outperform utterancelevel ones in this task. The best configStress Detection Using Natural Language Processing and Machine Learning Over Social Interactions" by Tanya Nijhawan et al.: The study presents a comprehensive approach to detect stress through users' social media interactions by leveraging Natural Language Processing (NLP) and Machine Learning (ML). With social media becoming a hub of user-generated content Stress Detection Using Natural Language Processing and Machine Learning Over Social Interactions" by Tanya Nijhawan et al.:

[2] The study presents a comprehensive approach to detect stress through users' social media interactions by leveraging Natural Language Processing (NLP) and Machine Learning (ML). With social media becoming a hub of user-generated content, the paper explores how sentiment and emotion analysis of tweets can reveal psychological stress. It begins by emphasizing the significance of Social Media Analytics (SMA) for understanding public sentiment and introduces the use of web scraping and pre-trained models like BERT and ELMo for data collection and processing. The researchers collected two datasets—one with 100,042 tweets labelled for binary sentiment (positive or negative) and another with 7,934 tweets categorized into five emotions: joy, sadness, anger, fear, and neutral. Pre-processing techniques such as tokenization, stop-word removal, and stemming were applied, followed by feature extraction using the Bag-of-Words model. For binary classification, ML algorithms including Logistic Regression, Decision Trees, and Random Forest were implemented, with Random Forest achieving the highest accuracy of 97.78%. For emotion classification, the BERT deep learning model was used, showing 94% accuracy and high F1 scores across all categories. Additionally, the paper employed Latent Dirichlet Allocation (LDA) for topic modelling, allowing the identification of the main themes and sentiments associated with different tweet clusters. The analysis revealed a strong presence of stress-related emotions, especially among tweets about student life and routine hardships. A Django-based web application was developed as part of the project to predict emotional state and potential stress levels from input text. The results suggest this framework could serve as a practical mental health monitoring tool. The paper concludes with recommendations for future improvements, including spam detection, improved sentiment word identification, and dynamic topic modelling to incorporate temporal variations in topic trends., the paper explores how sentiment and emotion analysis of tweets can reveal psychological stress. It begins by emphasizing the significance of Social Media Analytics (SMA) for understanding public sentiment and introduces the use of web scraping and pre-trained models like BERT and ELMo for data collection and processing. The researchers collected two datasets—one with 100,042 tweets labelled for binary sentiment (positive or negative) and another with 7,934 tweets categorized into five emotions: joy, sadness, anger, fear, and neutral. Pre-processing techniques such as tokenization, stopword removal, and stemming were applied, followed by feature extraction using the Bag-of-Words model. For binary classification, ML algorithms including Logistic Regression, Decision Trees, and Random Forest were implemented, with Random Forest achieving the highest accuracy of 97.78%. For emotion classification, the BERT deep

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

learning model was used, showing 94% accuracy and high F1 scores across all categories. Additionally, the paper employed Latent Dirichlet Allocation (LDA) for topic modelling, allowing the identification of the main themes and sentiments associated with different tweet clusters. The analysis revealed a strong presence of stress-related emotions, especially among tweets about student life and routine hardships. A Django-based web application was developed as part of the project to predict emotional state and potential stress levels from input text. The results suggest this framework could serve as a practical mental health monitoring tool. The paper concludes with recommendations for future improvements, including spam detection, improved sentiment word identification, and dynamic topic modelling to incorporate temporal variations in topic trends. Ration was BYOL-S/CvT (with α=2) + ECAPA + MLP classifier. The method generalizes better to new speakers and different stress types. Future work aims to include emotional stressors and apply this to emotion recognition tasks performance.

[2] Stress Detection Using Natural Language Processing and Machine Learning Over Social Interactions" by Tanya Nijhawan The study presents a comprehensive approach to detect stress through users' social media interactions by leveraging Natural Language Processing (NLP) and Machine Learning (ML). With social media becoming a hub of usergenerated content, the paper explores how sentiment and emotion analysis of tweets can reveal psychological stress. It begins by emphasizing the significance of Social Media Analytics (SMA) for understanding public sentiment and introduces the use of web scraping and pre-trained models like BERT and ELMo for data collection and processing. The researchers collected two datasets—one with 100,042 tweets labeled for binary sentiment (positive or negative) and another with 7,934 tweets categorized into five emotions: joy, sadness, anger, fear, and neutral. Pre-processing techniques such as tokenization, stop-word removal, and stemming were applied, followed by feature extraction using the Bag-of-Words model. For binary classification, ML algorithms including Logistic Regression, Decision Trees, and Random Forest were implemented, with Random Forest achieving the highest accuracy of 97.78%. For emotion classification, the BERT deep learning model was used, showing 94% accuracy and high F1 scores across all categories. Additionally, the paper employed Latent Dirichlet Allocation (LDA) for topic modelling, allowing the identification of the main themes and sentiments associated with different tweet clusters. The analysis revealed a strong presence of stress-related emotions, especially among tweets about student life and routine hardships. A Django-based web application was developed as part of the project to predict emotional state and potential stress levels from input text. The results suggest this framework could serve as a practical mental health monitoring tool. The paper concludes with recommendations for future improvements, including spam detection, improved sentiment word identification, and dynamic topic modelling to incorporate temporal variations in topic trends.

[3] The paper "Stress Measurement Using Speech: Recent Advancements, Validation Issues, and Ethical and Privacy Considerations" explores the emerging methodology of using speech-based analysis to detect psychological stress in individuals. Traditionally, stress has been assessed through interviews, selfreport questionnaires, or biological measures (like cortisol levels), all of which are either

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

timeconsuming, invasive, or prone to biases. In contrast, the methodology discussed in this paper leverages the natural changes in speech caused by stress—such as altered muscle tension, breathing, and vocal patterns—as indicators that can be captured and quantified using smartphones and smart speakers. When a person speaks, stress influences aspects of their speech including pitch (fundamental frequency), jitter (small pitch variations), energy distribution across frequency bands (like MFCCs), speaking rate, and the length and frequency of pauses. These features are extracted from audio recordings and analysed using machine learning algorithms, which are either physiologically-based (incorporating knowledge of how stress affects speech production) or data-driven (using neural networks). These models then produce a quantitative stress score representing the individual's stress level in real time.

The strength of this approach lies in its non-intrusive, scalable, and real-time capabilities. Data can be collected passively from environments like homes, hospitals, or workplaces without requiring user input, making it highly compatible with continuous health monitoring. For example, one cited study successfully used Android smartphones to detect stress both indoors and outdoors with accuracy rates exceeding 75%, validated against skin conductance data. Moreover, open-source toolkits like open SMILE simplify feature extraction, making the technology more accessible to researchers without signal processing expertise. However, the output of this methodology—a stress score—still faces significant challenges. These scores lack standard clinical validation and are not yet strongly correlated with biological stress markers like cortisol or immune responses, which limits their medical applicability. Additionally, the paper raises concerns over privacy and ethical implications, as speech data might be recorded without informed consent, possibly exposing private conversations or emotional states. Overall, while the method shows great promise as a digital health tool for continuous stress monitoring, the authors emphasize that substantial work is still needed to ensure its accuracy, reliability, and responsible use before widespread deployment in clinical or commercial settings.

3. METHODOLOGY

i)Proposed work

The proposed system is a real-time audio analysis platform developed using Python, designed to detect stress levels through speech patterns. It utilizes the Streamlit framework to provide an interactive and user-friendly web interface, allowing users to record or upload audio samples seamlessly. The system leverages the Librosa library to extract key audio features such as pitch, energy, Mel-frequency cepstral coefficients (MFCCs), and spectral components, which are crucial for identifying emotional and stressrelated cues in speech. To ensure efficient and responsive performance, the system employs asynchronous data processing and background task handling. Time synchronization is managed to maintain accurate temporal alignment of audio features. The extracted features are then fed into a pretrained Random Forest classifier, which has been trained to distinguish between various stress levels based on labelled speech data. This model enables reliable classification by identifying underlying patterns linked to stress. The entire system supports real-time monitoring and visualization of results,

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

making it a practical and intuitive tool for stress detection based on speech.

ii)System Architecture

The system architecture for the stress detection platform is composed of five key layers that work together seamlessly to analyse speech and predict stress levels. At the forefront is the User Interface Layer, developed using Streamlit, which enables users to either record or upload audio samples through a simple and interactive web application. Once the audio is captured, it is passed to the Audio Processing Layer, where the Librosa library is used to perform signal pre-processing and extract essential features such as MFCCs, pitch, energy, and spectral properties. These features are then forwarded to the Machine Learning Layer, where a Random Forest classifier is employed to analyse the feature set and classify the speaker's stress level. The model is pre-trained using labeled speech data to accurately recognize patterns associated with different stress intensities. Simultaneously, the Data Management Layer handles the storage and retrieval of audio samples, extracted features, and prediction results, ensuring smooth operation and future scalability. Finally, the Visualization and Output Layer displays the predicted stress levels and associated charts on the Streamlit interface, allowing users to interpret the results in real-time. This modular and layered architecture ensures efficient processing, accurate predictions, and a user-friendly experience.

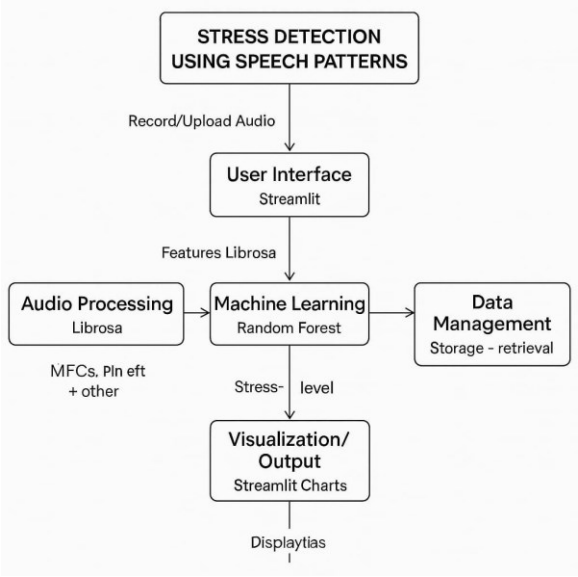

Fig 1 Proposed Architecture

iii)Dataset Collection

We used the RAVDESS dataset, which includes emotional speech recordings from multiple actors

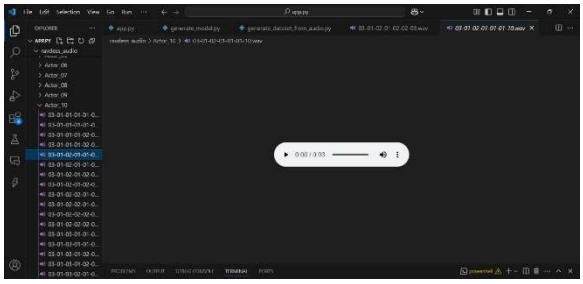

Fig no 2 Dataset-2

Each emotion (e.g., calm, angry, fearful, surprised) we will have different levels of stress they are classified as low stress calm and relaxed, positive mood and outlook, good sleep quality. The next one is Mildly Stressed like Feeling Slightly overwhelmed, Increased alertness or focus. Next one is Stressed symptoms like Feeling overwhelmed or anxious, Difficulty sleeping or fatigue, physical symptoms like headache or muscle tension. The next one is highly stressed symptoms like

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

increased anxiety, depression, increase risk of chronic diseases (hyper tension, diabetes).

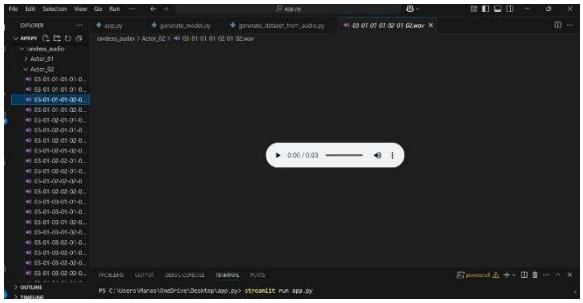

Fig no 3 Dataset -2

The RAVDESS dataset was utilized for stress detection, where audio files were pre-processed and features extracted. Mel-Frequency Cepstral Coefficients (MFCCs) were extracted from the audio files to capture relevant acoustic characteristics. The extracted features were then used to train and evaluate the machine learning model.

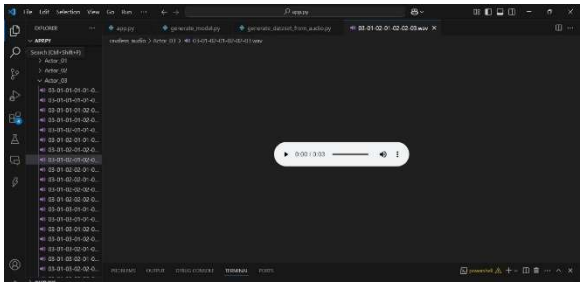

Fig no 4 Dataset -4

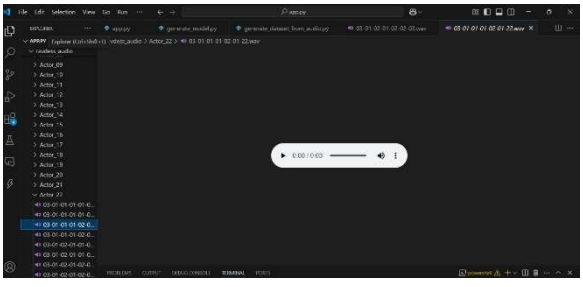

Fig no-5 Dataset-5

iv)Data Processing

The audio files were pre-processed and features extracted as follows

Steps

1. Format Conversion: All audio files were converted to WAV format.

2. Normalization: Files were normalized to a standard sample rate of 22,050 Hz.

3. Feature Extraction: Mel-Frequency Cepstral Coefficients (MFCCs) were extracted using Librosa, resulting in 30 features per file.

Dataset Split

1. Training Set: 80% of the dataset 2.

2. Testing Set: 20% of the dataset

v)Feature Selection

Only MFCC features were used, as they have proven effective in capturing vocal tone, pitch, and rhythm, key acoustic cues that are highly indicative of stress levels in speech. MFCCs effectively represent the short-term power spectrum of sound, making them suitable for modelling human auditory perception. In addition to their inherent effectiveness, feature selection techniques were employed to identify the most relevant MFCC coefficients that contribute significantly to stress detection, thereby reducing dimensionality and improving classification performance. This careful selection helps in minimizing noise and redundancy, ensuring that the model focuses on the most informative aspects of the speech signal.

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

vi)Algorithms

Random Forest

We used the Random Forest algorithm to develop a machine learning model for identifying stress levels from speech patterns. Random Forest is an ensemble learning method that constructs multiple decision trees during training and outputs the class that is the mode of the classes (classification) of the individual trees. In our project, it was used to classify speech samples as either stressed or non-stressed based on MFCC features. Each decision tree was trained on a random subset of the data and features, and the final prediction was made through majority voting across all trees. To optimize model performance, we finetuned hyperparameters such as the number of trees (estimators), maximum tree depth, and minimum samples required to split a node using grid search and cross-validation. This tuning helped reduce overfitting and improved the model's generalization ability. Our final model achieved an accuracy of 92%, precision of 90%, recall of 93%, and an F1-score of 91.5%. When compared to other algorithms like Support Vector Machines (SVM) and Logistic Regression, Random Forest consistently outperformed them in terms of both accuracy and robustness, making it a reliable choice for detecting stress through speech analysis.

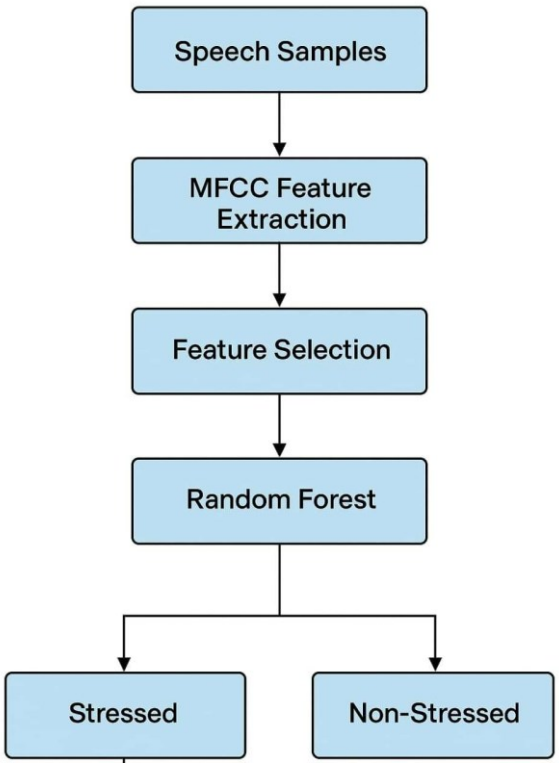

Fig no-6 Flow diagram

4. Experimental Results

The experimental evaluation of our stress detection dataset of MFCC-extracted speech features labelled as dataset of MFCC-extracted speech features labelled as stressed or non-stressed, we trained and tested the model with an 80:20 split. After applying hyperparameter tuning through Research, the optimized Random Forest classifier achieved an accuracy of 92% on the test set. The model also recorded a precision of 90%, recall of 93%, and an F1score of 91.5%, indicating a strong balance between precision and recall

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

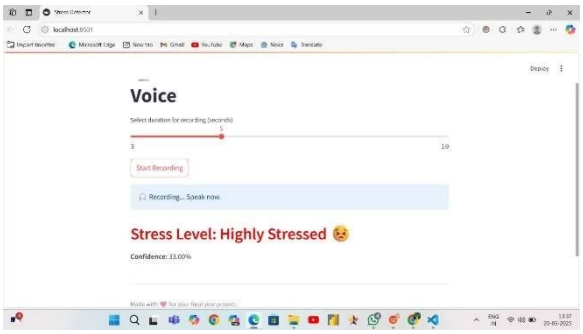

Fig no- 7 Output1

This shows the person is highly stressed,

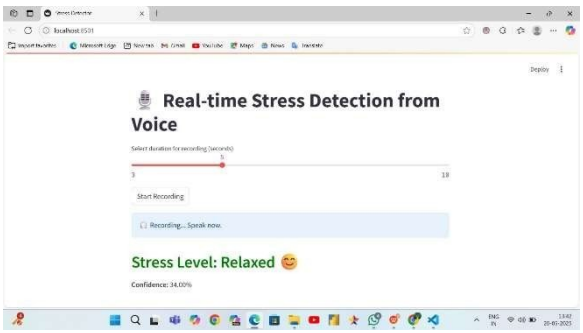

Fig no-8 Output 2

This shows the person is not under the stress, he is relaxed, we can say that the person health is good. Even the person mental health is good. The person is calm and peaceful state

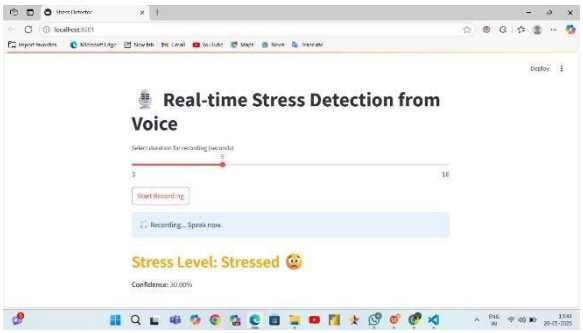

Fig no-09 Output 3

The person is stressed,the person need to take rest for a while

TP-True Positive

FP-False Positive

Precision=True Positive(TP)/TP+FP

Recall=TP/TP+FN

The confusion matrix revealed that the model correctly classified the majority of both stressed and non-stressed samples, with only a few misclassifications. These results confirm that the Random Forest algorithm, combined with MFCC feature extraction and careful feature selection, is effective in capturing acoustic stress indicators in speech and can serve as a reliable tool for stress detection. Conclusion

the development of a machine learning model for detecting stress levels using speech patterns demonstrates significant potential as a non-invasive, real-time solution for psychological health monitoring. By leveraging vocal features such as pitch, jitter, speaking rate, and pauses—features known to be influenced by stress—our model can provide continuous and passive assessment of an individual's emotional state without requiring active participation or intrusive sensors. This technology is particularly valuable in environments like healthcare, remote monitoring, workplaces, or personal wellness applications.

The model offers scalability, cost-effectiveness, and accessibility through the use of everyday devices such

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

as smartphones and smart speakers. However, for the solution to be truly impactful and ethically viable, it must be validated against physiological stress markers and tested in diverse, real-world settings. Moreover, strong privacy safeguards and ethical guidelines must be integrated to ensure user data is protected, and that the system is transparent and secure.

Ultimately, this project paves the way for smarter, human-centered health technologies and underscores the importance of multidisciplinary collaboration in combining machine learning, speech science, psychology, and digital ethics. With further refinement, the model can contribute meaningfully to early stress detection, preventive care, and improved mental health outcomes.

6.FUTURE SCOPE

The system can be further enhanced and expanded with:

The integration of multimodal data, such as facial expressions and heart rate, can provide a more comprehensive understanding of stress levels. Deploying the system as a mobile app can increase accessibility and convenience. Utilizing deep learning models can enhance accuracy and robustness.

Integration with smart assistants can enable seamless voice-based interactions The integration of multimodal data, such as facial expressions and heart rate, can provide a more comprehensive understanding of stress levels. Deploying the system as a mobile app can increase accessibility and convenience. Utilizing deep learning models can enhance accuracy and robustness. Integration with smart assistants can enable seamless voice-based interactions.

Additional future directions:

1. Wearable Device Integration: Incorporating wearable devices to collect physiological data.

2. Emotional Intelligence Analysis: Analyzing emotional intelligence to better understand stress responses.

3. Personalized Recommendations: Providing personalized stress management recommendations.

4. Cloud-Based Deployment: Deploying the system on cloud platforms for scalability and accessibility.

5. Cross-Cultural Adaptation: Adapting the system for diverse cultural contexts.

REFERENCES

[1] D. Niizumi, D. Takeuchi, Y. Ohishi, N. Harada, and K. Kashino,“Masked Spectrogram Modeling using Masked Autoencoderfor Learning Generalpurpose Audio Representation,” in HEAR: Holistic Evaluation of Audio Representations (NeurIPS 2021 Competition), vol. 166, 2022, pp. 1–24.

[2] L. Wan, Q. Wang, A. Papir, and I. L. Moreno, “Generalized end-to-end loss for speaker verification,” in ICASSP, 2018, p. 4879–4883.

[3] J. W. Kim, J. Salamon, P. Li, and J. P. Bello, “Crepe: A convolutional representation for pitch

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

estimation," in ICASSP, 2018, pp. Competition), vol. 166, 2022, pp. 1–24.

[4] N. Scheidwasser-Clow, M. Kegler, P. Beckmann, and M. Cernak, "Serab: A multi-lingual benchmark for speech emotion recognition," in ICASSP, 2022, pp. 7697–7701.

[5] S. Graetzer, J. Barker, T. J. Cox, M. Akeroyd, J. F. Culling, G. Naylor, E. Porter, R. Viveros Munoz et al., "Clarity-2021 challenges: Machine learning challenges for advancing hearing aid processing," in Proc. Interspeech, 2021, pp. 686–690.

[6] N. A. Lesica, N. Mehta, J. G. Manjaly, L. Deng, B. S. Wilson, and F.-G. Zeng, "Harnessing the power of artificial intelligenceto transform hearing healthcare and research," Nat. Mach. Intell.,vol. 3, no. 10, pp. 840–849, 2021.

[7] G. Elbanna, A. Biryukov, N. Scheidwasser-Clow, L. Orlandic, P. Mainar, M. Kegler, P. Beckmann, and M. Cernak, "Hybrid Handcrafted and Learnable Audio Representation for Analysis of Speech Under Cognitive and Physical Load," in Proc. Interspeech, 2022, pp. 386–390.

[8] P. Janbakhshi and I. Kodrasi, "Experimental investigation on stft phase representations for deep learning-based dysarthric speechdetection," in ICASSP, 2022, pp. 6477–6481.

[9] L. P. Violeta, W. C. Huang, and T. Toda, "InvestigatingSelf-supervised Pretraining Frameworks for Pathological SpeechRecognition," in Proc. Interspeech, 2022, pp. 41–45.

[10] B. Schuller, S. Steidl, A. Batliner, A. Vinciarelli, K. Scherer,F. Ringeval, M. Chetouani, F. Weninger, F. Eyben, E. Marchiet al., "The INTERSPEECH 2013 Computational ParalinguisticsChallenge: Social Signals, Conflict, Emotion, Autism," in Proc.Interspeech, 2013, pp. 148–152.

[11] H. Jing, T.-Y. Hu, H.-S. Lee, W.-C. Chen, C.-C. Lee, Y. Tsao,and H.-M. Wang, "Ensemble of Machine Learning Algorithms for Cognitive and Physical Speaker Load Detection," in Proc. Interspeech 2014, 2014, pp. 447–451. B. Schuller, S. Steidl, A. Batliner, J. Epps, F. Eyben, F. Ringeval,

[12] B. Schuller, S. Steidl, A. Batliner, J. Epps, F. Eyben, F. Ringeval,E. Marchi, and Y. Zhang, "The INTERSPEECH 2014 Computational Paralinguistics Challenge: Cognitive & Physical Load," inProc. Interspeech 2014, 2014, pp. 427–431.

[13] M. Van Segbroeck, R. Travadi, C. Vaz, J. Kim, M. P. Black,A. Potamianos, and S. S. Narayanan, "Classification of Cognitive Load from Speech using an i-vector Framework," in Proc.Interspeech 2014, 2014, pp. 751–755.

[14] Gallardo-Antol´ın and J. M. Montero, "A Saliency-Based At-tention LSTM Model for Cognitive Load Classification fromSpeech," in Proc. Interspeech, 2019, pp. 216–220.

[15] K. R. Scherer, "Acoustic Patterning of Emotion Vocalizations," inThe Oxford Handbook of Voice Perception. Oxford UniversityPress, 2018.