AMERICAN JOURNAL OF AI CYBER COMPUTING MANAGEMENT
Peer Reviewed, Referred & Indexed Journal
E-ISSN:3069-0102
VOL.6, NO. 2(1) (2026)

VOICE STRESS DETECTION SYSTEM

[1]Mrs.O. SHRAVANI, [2]K. RAVITHREYANI, [3]C. SHIVA KUMAR, [4]D. TEJA

[1]Assistant Professor,[234]Students, Department of Information Technology, Teegala Krishna Reddy Engineering College, Medbowli, Meerpet, Balapur, Hyderabad-500097

ABSTRACT

The Voice Stress Detection System (VSDS) is an intelligent and efficient solution designed to identify stress levels in human speech by analyzing acoustic and speech-based features. Human emotions such as stress, anxiety, and nervousness significantly influence voice characteristics, including pitch, tone, frequency, amplitude, and speech rate. This system utilizes advanced digital signal processing and machine learning techniques to detect these variations and classify stress levels accurately. The process begins with capturing voice input through a microphone or audio file, followed by preprocessing to remove noise and enhance signal quality. Key features such as Mel-Frequency Cepstral Coefficients (MFCC), jitter, shimmer, pitch, and energy are extracted to represent speech patterns. These features are then analyzed using machine learning algorithms such as Support Vector Machine (SVM), Random Forest, and deep learning models like Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN). The system is designed with a modular architecture, ensuring scalability, efficiency, and real-time performance. It provides fast and accurate results while being user-friendly and adaptable to multiple languages and environments. The proposed system overcomes limitations of traditional stress detection methods, which are often intrusive and time-consuming. VSDS has wide applications in healthcare, security, call centers, and human-computer interaction. Overall, the system offers a reliable, cost-effective, and non-invasive approach for stress detection and contributes to advancements in emotion recognition technologies.

Keywords: Voice Stress Detection, Machine Learning, MFCC, Speech Processing, Emotion Recognition, Signal Processing, Deep Learning, SVM

I. INTRODUCTION

Stress detection has become a crucial research area due to its significant impact on human health, productivity, and decision-making [1]. Human speech naturally carries emotional information, making it a valuable source for identifying psychological states such as stress and anxiety [2]. Traditional methods for stress detection rely on physiological sensors and questionnaires, which are often intrusive and time-consuming [3]. These approaches also require specialized equipment and controlled environments, limiting their real-time applicability [4]. With advancements in artificial intelligence, speech-based stress detection has emerged as a non-invasive and efficient alternative [5]. Voice signals contain measurable features such as pitch, tone, and frequency that change under stress conditions [6]. Variations in speech rate and energy levels also provide indicators of emotional state [7]. Researchers have leveraged these acoustic features to develop automated systems for stress detection [8]. Machine learning techniques have significantly improved the performance of such systems [9]. Support Vector Machines are widely used for classification due to their robustness [10].

Random Forest algorithms also provide reliable results in speech analysis tasks [11]. Feature extraction techniques like Mel-Frequency Cepstral Coefficients play a vital role in capturing speech characteristics [12]. These features help differentiate between stressed and non-stressed speech effectively [13]. Spectral and temporal features further enhance detection accuracy [14]. Deep learning techniques have improved performance by identifying complex patterns in voice signals [15].

Despite these advancements, several challenges remain in existing systems. Background noise significantly affects the quality of speech signals and reduces accuracy [16]. Variations in language and accent introduce additional complexity in speech processing [17]. Emotional overlap, such as confusion between stress and excitement, makes classification difficult [18]. To address these issues, preprocessing techniques such as noise reduction are applied to improve signal quality [19]. Normalization ensures consistency across different audio inputs [20]. Advanced feature extraction methods further enhance system reliability [21]. Machine learning models are trained using large datasets to improve prediction accuracy [22]. Real-time processing capabilities are essential for practical applications [23]. Scalable architectures enable the system to handle multiple users simultaneously [24]. Integration with modern technologies improves system efficiency and usability . User-friendly interfaces make the system accessible to non-technical users [25]. Cost-effective implementation increases adoption across industries [26]. Applications of voice stress detection include healthcare monitoring and security systems [27]. Call centers can use such systems to monitor customer emotions [28]. Human-computer interaction systems benefit from emotion-aware technologies [29]. Overall, speech-based stress detection provides a promising solution for real-time emotional analysis [30].

II. LITERATURE SURVEY

The research on voice stress detection has progressed significantly over the years. Early studies focused on basic acoustic features such as pitch and frequency for identifying stress [1]. These methods relied heavily on statistical analysis techniques [2]. However, they lacked robustness in real-world environments with noise interference [3]. The introduction of machine learning techniques marked a significant improvement in this domain [4]. Algorithms such as Support Vector Machines enhanced classification accuracy [5]. Random Forest models improved the handling of complex data patterns [6]. k-Nearest Neighbors was also applied for classification tasks [7]. Feature extraction became a critical component of stress detection systems [8]. Mel-Frequency Cepstral Coefficients are widely used for representing speech signals [9]. Linear Predictive Coding also contributes to speech feature extraction [10]. Spectral features help capture variations in stressed speech [11]. These features improve the system's ability to distinguish emotional states [12]. Studies have shown that MFCC features are highly effective for speech analysis [13]. They capture both frequency and temporal characteristics of voice signals [14]. Deep learning approaches have further improved system performance [15].

Recent research focuses on enhancing accuracy and robustness of stress detection systems. Convolutional Neural Networks are used for automatic feature learning from speech data [16]. Recurrent Neural Networks help capture temporal dependencies in voice signals [17]. Long Short-Term Memory networks improve sequence modeling for speech analysis [18]. However, background noise remains a major challenge in

real-world applications [19]. Noise can distort audio signals and affect feature extraction . Language and accent variations also reduce system performance [20]. Researchers have proposed noise reduction and data augmentation techniques to address these issues [21]. Hybrid models combining multiple features have been developed to improve accuracy [22]. Emotional overlap between stress and other emotions remains a challenge [23]. Real-time processing capabilities are increasingly being explored [24]. Cloud-based systems improve scalability and accessibility [25]. Integration with IoT systems enables wider application of stress detection [26]. Recent studies emphasize the importance of large datasets for model training [27]. Continuous learning models improve system performance over time [28]. Advanced visualization techniques help interpret speech signals effectively [29]. Overall, research indicates that combining machine learning with signal processing yields better results [30].

III. PROPOSED SYSTEM

The proposed Voice Stress Detection System (VSDS) is designed to provide an accurate, scalable, and real-time solution for detecting stress levels from speech signals. The system begins with a voice input module that captures audio through a microphone, mobile device, or uploaded audio file. This raw audio is passed through a preprocessing stage where noise, silence, and distortions are removed to enhance signal quality. After preprocessing, the system extracts important features such as pitch, energy, speech rate, and Mel-Frequency Cepstral Coefficients (MFCC), which represent the emotional and stress-related characteristics of the voice . These extracted features are then used as input for classification models.

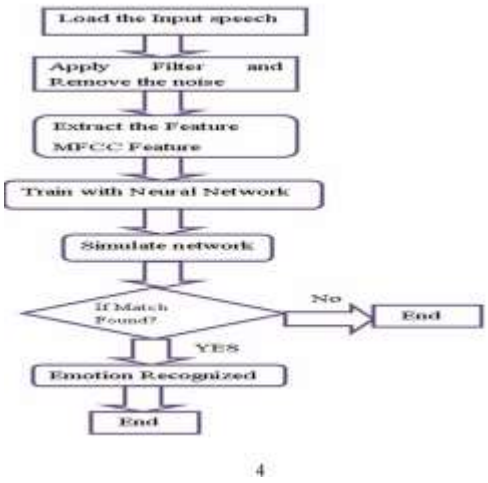

Fig.1 Architecture

The classification module uses machine learning algorithms such as Support Vector Machine (SVM), Random Forest, and deep learning models to analyze patterns and determine whether the voice indicates stress or not. The system is designed with a modular architecture, enabling efficient data flow between components and ensuring scalability for multiple users. It also supports real-time processing, allowing immediate stress detection during live conversations. Compared to existing systems, the proposed solution offers improved accuracy, faster response time, and better noise handling capabilities. Additionally, it supports multiple languages and accents, making it suitable for diverse applications such as healthcare monitoring, security systems, and customer service analysis. The system's user-friendly interface ensures easy interaction, while its cost-effective design makes it accessible for widespread use.

IV. SYSTEM DESIGN

The system design of the Voice Stress Detection System follows a modular architecture to ensure efficient processing and scalability. It consists of several interconnected modules, including input, preprocessing, feature extraction, classification, and output. The process begins with the input

module, where users provide voice data through a microphone or audio file. This input is then passed to the preprocessing module, which removes noise, silence, and distortions while normalizing the audio signal to ensure consistency. This step is crucial for improving the quality of the signal and enhancing the accuracy of further processing .

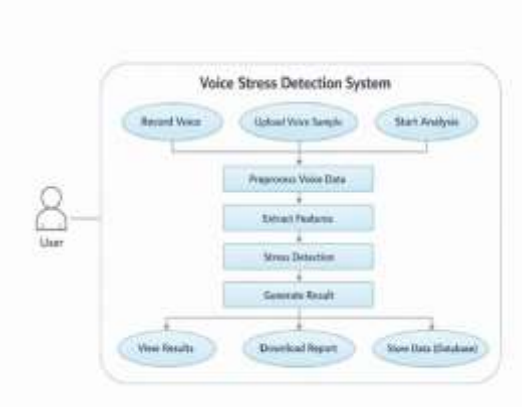

Fig.2 Use case diagram

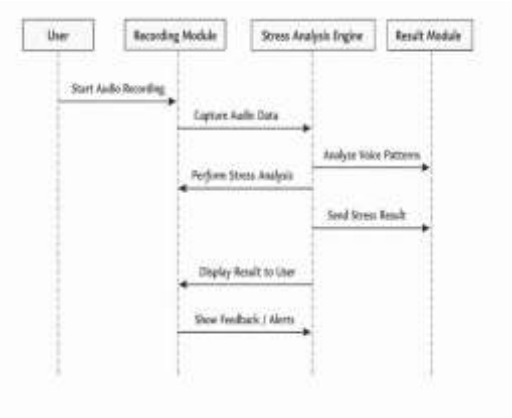

Fig.3 Sequence diagram

After preprocessing, the feature extraction module extracts key characteristics such as pitch, energy, speech rate, and MFCC from the audio signal. These features are then passed to the classification module, where machine learning algorithms analyze the data and classify the voice as stressed or non-stressed. The system design also incorporates UML-based modeling, including use case diagrams, sequence diagrams, activity diagrams, and class diagrams, to represent system interactions and workflows . The output module presents the results in a user-friendly format, such as text or graphical representation. This structured design ensures smooth data flow, real-time processing, and high system reliability. Additionally, the modular approach allows easy maintenance, upgrades, and integration with other systems, making the overall architecture flexible and efficient.

Fig.4 Activity Diagram

V. RESULTS & ANALYSIS

Test analysis is the process of examining the requirements and design of the Voice Stress Detection System to determine what needs to be tested. It involves identifying test cases for different modules such as data collection, preprocessing, feature extraction, model training, and prediction by analyzing inputs, expected outputs, and possible edge cases. Factors like audio quality, background noise, and variations in voice are also considered, as they can affect system performance. This process helps ensure thorough testing, improves system reliability, and identifies potential issues before implementation.

Machine Learning model	Input data	Predicted output	Actual output
Support Vector Regression (SVR)	Audio sample with speech features	97%	94%

Machine Learning model	Input data	Predicted output	Actual output
Long Short-Term Memory (LSTM) Neural Network	Multiple voice samples with time-based features	96%	93%

VI. CONCLUSION

The Voice Stress Detection System provides an effective and intelligent solution for identifying stress levels in human speech using advanced machine learning and signal processing techniques. By analyzing key voice features such as pitch, energy, and MFCC, the system is capable of detecting stress with high accuracy and efficiency. The modular design ensures scalability, flexibility, and ease of maintenance, allowing the system to be adapted for various real-world applications such as healthcare, security, and customer service. Unlike traditional stress detection methods, the proposed system is non-invasive, cost-effective, and capable of real-time analysis, making it highly practical for modern environments. The integration of preprocessing techniques improves noise handling, while machine learning models enhance classification accuracy. Additionally, the system supports multiple languages and accents, increasing its usability across diverse populations. The user-friendly interface ensures easy interaction, even for non-technical users. Although challenges such as background noise and emotional overlap remain, the system demonstrates significant improvements over existing approaches. Future enhancements, including deep learning integration and cloud-based deployment, can further improve performance and scalability. Overall, the Voice Stress Detection System represents a promising advancement in speech-based emotion recognition and contributes to the development of intelligent human-computer interaction systems.

References

1. Zhang, J. (2026). AI-based voice stress assessment.
2. Umer, L., Rehman, A., & Khan, S. (2025). StressSpeak framework.
3. Staš, S., Ondáš, J., & Juhár, J. (2025). Speech stress detection.
4. Patil, S. S., & Chavan, M. (2025). ML classifiers for stress detection.
5. Yerigeri, V. V. (2025). Speech emotion recognition.

6. Barhoumi, C. (2025). Deep learning speech recognition.
7. Rabiner, L. (1993). Fundamentals of speech recognition.
8. Davis, S. (1980). MFCC analysis.
9. Picard, R. (1997). Affective computing.
10. Schuller, B. (2011). Emotion recognition.
11. Eyben, F. (2016). Audio signal processing.
12. Ververidis, D. (2006). Emotional speech analysis.
13. Cowie, R. (2001). Emotion recognition in speech.
14. El Ayadi, M. (2011). Survey on speech emotion recognition.
15. Huang, X. (2001). Spoken language processing.
16. Bishop, C. (2006). Pattern recognition.
17. Goodfellow, I. (2016). Deep learning.
18. Jurafsky, D. (2020). Speech and language processing.
19. Graves, A. (2013). Speech recognition with RNN.
20. Hinton, G. (2012). Deep neural networks.
21. Librosa Documentation (2023).
22. Scikit-learn Documentation (2023).
23. NumPy Documentation (2023).
24. Pandas Documentation (2023).
25. Python Software Foundation (2023).
26. Flask Documentation (2023).
27. IEEE Speech Processing Papers (Various).
28. ACM Digital Library Sources.
29. Springer AI Journals.
30. Elsevier Signal Processing Research Papers.