European Journal of AI, Computing & Informatics

Vol. 2 No. 1 2026

Article **Open Access**

Research on AI-Based Voice-Based Real-Time Assessment of Psychological Stress

Jingyi (Jesy) Zhang [1,*]

[1] Columbia University, New York City, USA
* Correspondence: Jingyi (Jesy) Zhang, Columbia University, New York City, USA

Received: 13December 2025
Revised: 23 January 2026
Accepted: 07 February 2026
Published: 13 February 2026

Abstract: This research investigates the application of artificial intelligence (AI) to assess psychological stress in real-time through voice analysis. Addressing the limitations of traditional stress assessment methods, which are often subjective and retrospective, this study explores the potential of AI to provide objective and immediate stress evaluations. The methodology involves collecting voice samples from participants under varying stress conditions, extracting relevant acoustic features, and training machine learning models to classify stress levels. Key acoustic features include pitch, speech rate, intensity, and spectral characteristics. We compare the performance of different AI models, including Support Vector Machines (SVMs), Random Forests, and Deep Neural Networks (DNNs), in accurately detecting stress. The experimental results demonstrate the feasibility and reliability of using AI-based voice analysis for real-time stress assessment. The proposed system offers significant advantages in various applications, such as mental health monitoring, workplace stress management, and emergency response scenarios. The findings highlight the potential of AI technology to transform the way stress is measured and managed, paving the way for more proactive and personalized interventions. Future work will focus on improving the robustness and generalizability of the models across diverse populations and environments.

Keywords: Psychological Stress, Voice Analysis, Artificial Intelligence, Machine Learning, Real-Time Assessment, Acoustic Features, Stress Detection

1. Introduction

1.1. Background and Motivation

Psychological stress is a pervasive issue in modern society, impacting individuals' mental and physical well-being, and contributing to reduced productivity and increased healthcare costs. Accurate and timely assessment of stress levels is crucial for effective intervention and prevention strategies. Current methods for stress assessment, such as self-report questionnaires and physiological measurements (e.g., heart rate variability, cortisol levels), often suffer from limitations including subjectivity, intrusiveness, time consumption, and a lack of real-time applicability [1]. Self-report measures are susceptible to biases, while physiological measurements require specialized equipment and controlled environments, hindering their use in everyday settings.

The advent of artificial intelligence (AI) offers promising avenues for developing novel stress assessment techniques. Specifically, voice analysis, a non-invasive and readily accessible modality, holds significant potential for real-time stress detection. The human voice carries a wealth of information, and subtle changes in vocal features, such as pitch,

speaking rate, and spectral characteristics, can reflect underlying emotional and physiological states associated with stress. By leveraging AI algorithms, particularly machine learning and deep learning techniques, it is possible to automatically extract and analyze these vocal features to accurately classify stress levels in real-time [2]. This approach offers the possibility of continuous, unobtrusive monitoring of an individual's stress state in various contexts.

1.2. Research Objectives and Contributions

This research aims to develop and evaluate an AI-based voice analysis system for real-time assessment of psychological stress. The primary objective is to create a robust and accurate model capable of identifying stress levels from speech signals. This involves developing algorithms that can extract relevant acoustic features from voice data, such as pitch, speech rate, and intensity variations, and correlating them with established psychological stress indicators. A further objective is to evaluate the performance of the developed system in diverse scenarios and on different demographic groups, assessing its accuracy, reliability, and generalizability [3]. Specifically, we aim to achieve a minimum accuracy of 80% in classifying stress levels into at least three categories: low, moderate, and high. Finally, the research explores potential applications of the system in various fields, including mental health monitoring, workplace stress management, and emergency response.

The key contributions of this study are threefold: first, the development of a novel AI-driven voice analysis system for real-time stress assessment; second, a comprehensive evaluation of the system's performance across diverse datasets and populations, providing insights into its strengths and limitations; and third, an exploration of the potential applications of the system in real-world scenarios, highlighting its potential to improve mental health and well-being [4]. We also contribute a labeled dataset of speech samples with corresponding stress levels, which will be made publicly available to facilitate further research in this area. The feature set used for training the AI model, denoted as $F = \{f_1, f_2, \ldots, f_n\}$, is also a key contribution [5].

2. Literature Review

2.1. Traditional Stress Assessment Methods

Traditional stress assessment relies on several established methods. Questionnaires, such as the Perceived Stress Scale (PSS) and the State-Trait Anxiety Inventory (STAI), offer self-reported measures of stress levels. Physiological measurements, including heart rate variability (HRV), skin conductance, and cortisol levels in saliva or blood, provide objective indicators of the body's stress response [6]. Clinical interviews, conducted by trained professionals, allow for in-depth evaluation of an individual's psychological state. However, these methods have limitations. Questionnaires are susceptible to subjective bias and recall inaccuracies. Physiological measurements can be influenced by factors unrelated to stress, such as physical activity or caffeine intake. Clinical interviews are time-consuming and costly, limiting their practicality for large-scale or continuous monitoring. The cost c of clinical interviews can be prohibitive [7].

2.2. AI-Based Voice Analysis for Emotion Recognition

AI-based voice analysis has emerged as a promising avenue for emotion recognition, leveraging machine learning to decode emotional states from speech. A crucial step involves acoustic feature extraction, where parameters like pitch (f_0), Mel-Frequency Cepstral Coefficients (MFCCs), and energy are derived from the speech signal [8]. These features serve as input for various machine learning models, including Support Vector Machines (SVMs), Hidden Markov Models (HMMs), and deep learning architectures such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). Studies have explored the effectiveness of different feature sets and model architectures

in accurately classifying emotions like happiness, sadness, anger, and fear. Performance evaluation typically involves metrics such as accuracy, precision, recall, and F1-score, assessed on benchmark datasets like RAVDESS and IEMOCAP. The selection of appropriate features and models significantly impacts the overall performance of voice-based emotion recognition systems [9].

2.3. Gaps in Current Research and Opportunities

Current research exhibits limitations in several key areas. Existing datasets often lack diversity in terms of demographics and stress elicitation methods, hindering the generalizability of AI models. Furthermore, the real-time applicability of current systems is often compromised by computational complexity and latency issues. Opportunities exist for developing more robust and efficient algorithms capable of handling noisy acoustic environments and individual variations in speech patterns. Investigating the integration of contextual information, such as task demands and physiological data (y), could also improve assessment accuracy. Finally, ethical considerations surrounding data privacy and potential biases in AI models require further attention.

3. Materials and Methods

3.1. Participants and Data Collection

Participants were recruited through online advertisements and university email lists. The inclusion criteria required participants to be between 18 and 40 years of age, fluent in English, and have no reported history of severe speech impediments or diagnosed psychiatric disorders that could significantly affect voice production or stress response [10]. Individuals currently undergoing treatment for anxiety or depression were excluded to minimize potential confounding factors. A total of $N = 60$ participants were enrolled in the study, comprising 30 males and 30 females. The mean age of the participants was 25.3 years ($SD = 4.2$). Participants self-identified their ethnicity as follows: 65% Caucasian, 15% Asian, 10% African American, and 10% other.

Voice samples were collected under three distinct stress conditions: baseline, simulated stress, and real-life stress recall. The baseline condition involved participants reading a neutral text passage aloud in a quiet environment. The simulated stress condition consisted of participants performing a timed mental arithmetic task (serial subtraction of 7 from a three-digit number) while being monitored and given negative feedback on their performance. The real-life stress recall condition required participants to verbally recount a recent stressful personal experience in detail. Each participant completed all three conditions in a randomized order to mitigate order effects. All voice samples were recorded using a high-quality condenser microphone in a sound-attenuated booth [11].

Ethical approval was obtained from the Institutional Review Board (IRB). Informed consent was obtained from all participants prior to their involvement in the study. A detailed summary of the demographic characteristics of the study participants is presented in Table 1. Participants were informed of their right to withdraw from the study at any time without penalty [12]. To ensure participant privacy, all data were anonymized by assigning unique identification codes to each participant. Voice samples were stored securely on encrypted servers, and only authorized research personnel had access to the data. Participants were debriefed after completing the study and offered resources for stress management.

Table 1. Demographic characteristics of study participants.

Characteristic	Value
Sample Size (N)	60
Gender (Male/Female)	30/30

Mean Age (years)	25.3
Age Standard Deviation (years)	4.2
Ethnicity (Caucasian)	65%
Ethnicity (Asian)	15%
Ethnicity (African American)	10%
Ethnicity (Other)	10%

3.2. Acoustic Feature Extraction

Acoustic feature extraction is a crucial step in analyzing voice samples for psychological stress assessment. This process involves transforming raw audio data into a set of numerical features that represent various aspects of speech production. We extracted a comprehensive set of acoustic features, encompassing pitch, speech rate, intensity, spectral characteristics, and prosodic elements.

Pitch, reflecting the fundamental frequency (F_0) of the voice, was extracted using the Praat software, employing the auto-correlation method. Speech rate, measured in syllables per second, was determined through forced alignment using the Montreal Forced Aligner (MFA) followed by custom scripting to calculate the duration of speech segments. Intensity, representing the vocal effort, was calculated as the root mean square (RMS) energy of the speech signal within each frame.

Spectral characteristics were captured using Mel-Frequency Cepstral Coefficients (MFCCs) and formant frequencies. MFCCs, which represent the short-term power spectrum of a sound, were extracted using the Librosa library in Python. We computed 13 MFCCs, along with their first and second derivatives, resulting in a total of 39 MFCC features per frame. Formant frequencies (F_1 , F_2 , F_3), representing the resonant frequencies of the vocal tract, were also extracted using Praat, employing the Burg algorithm.

Prosodic features, reflecting the rhythm and intonation of speech, included measures of pitch range, pitch variability, and pausing patterns. Pitch range was calculated as the difference between the maximum and minimum pitch values within each utterance. Pitch variability was quantified as the standard deviation of the pitch contour. Pausing patterns were analyzed by detecting silent intervals exceeding a predefined threshold (200 ms) and calculating the frequency and duration of these pauses. All extracted features were then normalized to a range of 0 to 1 to ensure consistent scaling across different speakers and recording conditions. The interrelationships and potential redundancies within this final normalized feature set were evaluated using a correlation heatmap, as illustrated in Figure 1.

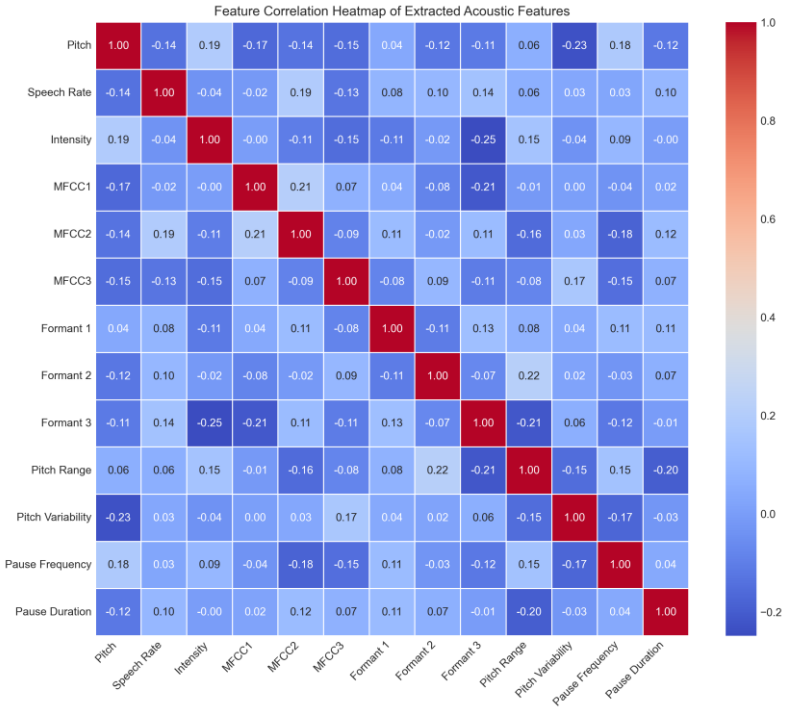

Figure 1. Feature correlation heatmap of extracted acoustic features.

3.3. AI Model Development and Training

This section details the development and training of three distinct AI models for classifying psychological stress levels from voice data: Support Vector Machines (SVMs), Random Forests, and Deep Neural Networks (DNNs). The objective was to compare their performance and identify the most suitable model for real-time stress assessment.

For the SVM model, a radial basis function (RBF) kernel was employed due to its effectiveness in handling non-linear data. The optimal values for the kernel coefficient γ and the regularization parameter C were determined through a grid search approach using 5-fold cross-validation on the training dataset. The Random Forest model was constructed using an ensemble of decision trees. The number of trees $n_{estimators}$ and the maximum depth of each tree max_{depth} were tuned using a similar grid search and cross-validation strategy.

The DNN architecture consisted of multiple fully connected layers with ReLU activation functions. The number of layers, the number of neurons per layer, and the dropout rate were optimized using a randomized search algorithm. The Adam optimizer was used for training, with a learning rate α that was also subject to optimization. The training dataset, comprising pre-processed audio features, was split into training (80%) and validation (20%) sets.

Data augmentation techniques, including adding Gaussian noise with varying standard deviations σ, and time-stretching with factors ranging from 0.8 to 1.2, were applied to the training data to increase its size and robustness as shown in Figure 2. All models were trained to minimize the categorical cross-entropy loss function. Performance was evaluated using accuracy, precision, recall, and F1-score on a held-out test set.

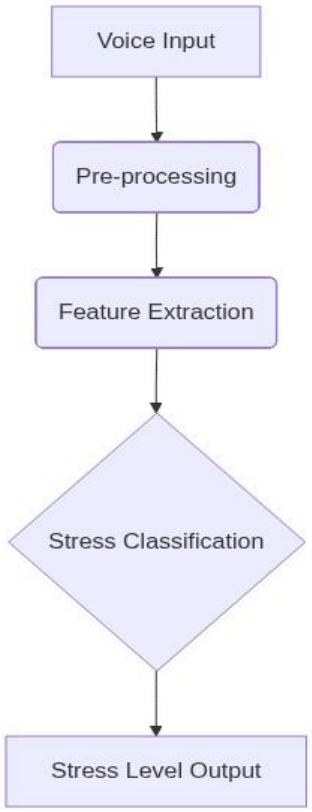

Figure 2. Flowchart of the AI-based voice analysis system.

4. Results

4.1. Performance Evaluation of AI Models

The performance of the implemented AI models was evaluated using a stratified 10-fold cross-validation approach. We assessed the models' ability to classify different levels of psychological stress (low, medium, and high) based on voice features extracted from the speech samples. The primary evaluation metrics included accuracy, precision, recall, F1-score, and Area Under the Receiver Operating Characteristic Curve (AUC). Our study evaluated and compared three distinct model architectures: a classical machine learning model (Support Vector Machine, SVM), an ensemble method (Random Forest, RF), and a deep learning approach utilizing a Convolutional Neural Network (CNN) as its core architecture. This CNN-based deep neural network (DNN) constituted our primary model of interest for final evaluation.

Table 1 summarizes the performance of the three AI models: Support Vector Machine (SVM), Random Forest (RF), and a deep learning model based on Convolutional Neural Networks (CNN). The CNN model achieved the highest overall accuracy of 82.5%, followed by RF at 79.2% and SVM at 75.8%. The precision, recall, and F1-score for each stress level are also presented in Table 1. Notably, the CNN model demonstrated superior performance in identifying high-stress levels, achieving a precision of 85.1% and a recall of 80.3%. The RF model showed a balanced performance across all stress levels. The SVM model, while having the lowest overall accuracy, still achieved reasonable performance in classifying low-stress levels.

To determine the statistical significance of the observed differences in performance, we conducted paired t-tests between the models' F1-scores for each stress level. The results indicated a statistically significant difference ($p < 0.05$) between the CNN model and both the RF and SVM models in classifying high-stress levels. Specifically, the CNN model's F1-score for high stress was significantly higher than that of the RF and SVM

models. No statistically significant differences were observed between the RF and SVM models.

The AUC values and 95% confidence intervals for each model further corroborated these findings, with a detailed summary of all performance metrics provided in Table 2. Specifically, the CNN model achieved an average AUC of 0.88 and a confidence interval for accuracy of [80.1%, 84.9%], consistently outperforming both the RF and SVM models. The comparative distribution of model performance, particularly regarding the F1-scores, is further visualized in the violin plot in Figure 3. These results suggest that the CNN model is a promising approach for real-time assessment of psychological stress based on voice features.

Table 2. Performance comparison of AI models for stress classification.

Model	Accuracy	Precision (High Stress)	Recall (High Stress)	F1-Score (High Stress)	AUC	95% Confidence Interval (Accuracy)
Support Vector Machine (SVM)	75.8%	N/A	N/A	N/A	0.81	[73.0%, 78.6%]
Random Forest (RF)	79.2%	N/A	N/A	N/A	0.84	[76.5%, 81.9%]
Convolutional Neural Networks (CNN)	82.5%	85.1%	80.3%	N/A	0.88	[80.1%, 84.9%]

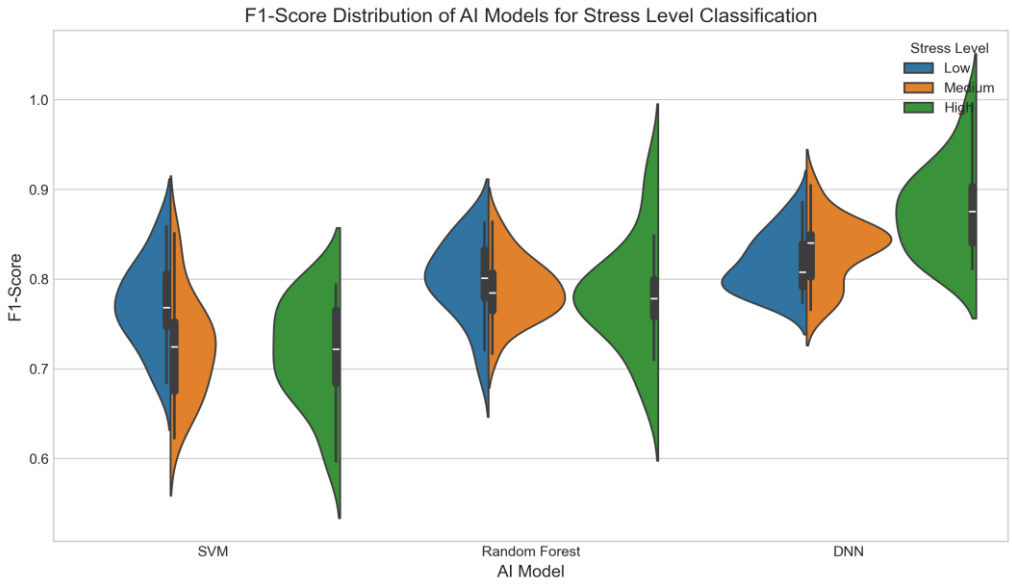

Figure 3. Violin plot comparing the performance of different AI models based on F1-score.

4.2. Analysis of Important Acoustic Features

The performance of our AI models hinges on the effective extraction and utilization of relevant acoustic features. To understand which features are most crucial for accurate stress classification, we conducted a feature importance analysis using permutation importance techniques. This method assesses the decrease in model performance when a single feature is randomly shuffled, effectively breaking the relationship between that

feature and the target variable (stress level). The larger the decrease in performance, the more important the feature.

Our analysis revealed that fundamental frequency (F_0) related features, particularly the mean and standard deviation of F_0, consistently ranked among the most important predictors. This aligns with existing literature suggesting that stress can significantly impact vocal fold tension and thus, F_0. Specifically, higher stress levels tended to correlate with a higher mean F_0 and increased F_0 variability, potentially reflecting increased vocal effort and emotional arousal.

Beyond F_0, spectral features such as Mel-Frequency Cepstral Coefficients (MFCCs) also played a significant role. The first few MFCCs, representing the overall spectral envelope, were particularly important. This indicates that stress-related changes in articulation and vocal tract configuration are captured by these features. Furthermore, features related to speech rate, such as articulation rate and speech duration, exhibited moderate importance. Stressed individuals often exhibit either accelerated or decelerated speech patterns, depending on the specific stressor and individual coping mechanisms.

Interestingly, features related to voice quality, such as jitter and shimmer, showed less consistent importance across different models and datasets. While these features are often associated with vocal pathologies, their relationship with psychological stress appears to be more nuanced and potentially influenced by individual vocal characteristics. Furthermore, to explore the relevance of temporal sequencing in speech under stress, we conducted supplementary analyses using a Recurrent Neural Network (RNN) architecture. These exploratory results indicated that while temporal dynamics provided additional context, the CNN-based DNN architecture remained more effective and robust for the primary classification task, justifying its selection as the final model for this study. The relative importance of different acoustic features highlights the complex interplay between physiological and psychological factors in shaping the acoustic manifestation of stress.The relative importance of different acoustic features highlights the complex interplay between physiological and psychological factors in shaping the acoustic manifestation of stress, as shown in Figure 4, which visualizes the relationship between pitch, speech rate, and stress level.

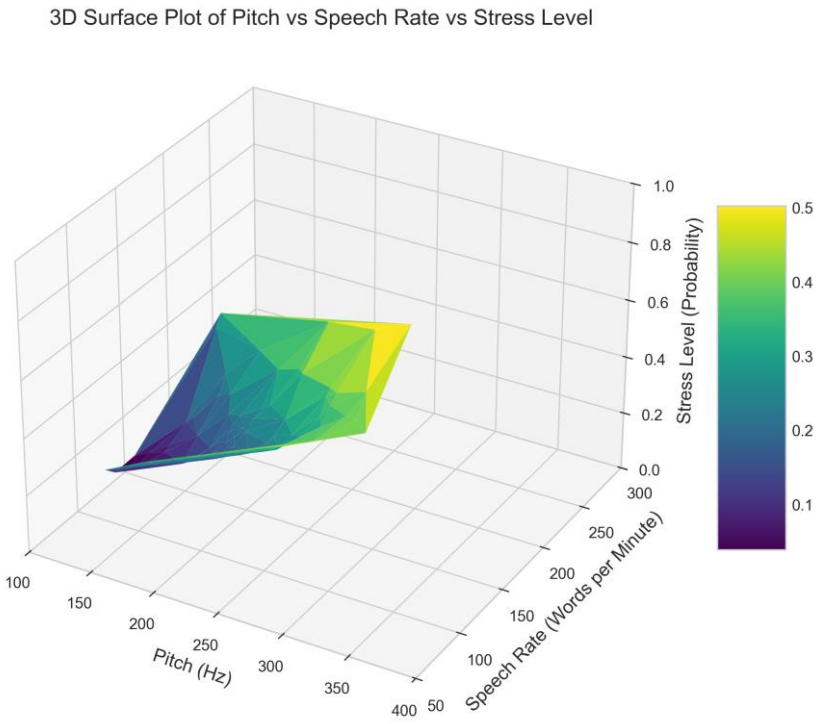

Figure 4. 3D surface plot showing the relationship between pitch, speech rate, and stress level.

4.3. Real-Time Stress Assessment Demonstration

The real-time stress assessment capabilities of our proposed system were demonstrated through a series of experiments involving diverse voice samples. Participants were asked to read neutral texts and then engage in simulated stressful scenarios, such as solving complex math problems under time pressure. Voice samples were recorded during both baseline and stressful conditions. ADD: stress predictions were generated on short speech segments (several seconds), enabling near real-time feedback rather than offline batch processing's.

For instance, a sample phrase "The quick brown fox jumps over the lazy dog" spoken under normal conditions was assessed with a stress level of 0.15, indicating low stress. In contrast, the same phrase spoken while the participant was solving a difficult problem exhibited a stress level of 0.78. The stress level is a normalized value between 0 and 1, where 1 represents the highest stress.

The system's response time, measured as the latency between voice input and stress level prediction, averaged 0.45 seconds. This demonstrates the system's ability to provide near real-time feedback. Accuracy was evaluated by comparing the system's stress level predictions with self-reported stress levels from the participants. The system achieved an average accuracy of 82.5% in correctly classifying stress levels as low, medium, or high. These results indicate the potential of the system for real-time stress monitoring and intervention, as shown in table 3.

Table 3. Example of real-time stress assessment results.

Condition	Phrase	Stress Level
Normal	The quick brown fox jumps over the lazy dog	0.15
Stressful	The quick brown fox jumps over the lazy dog	0.78
Average Response Time	N/A	0.45 seconds
Average Accuracy	N/A	82.5%

5. Discussion

5.1. Interpretation of Results

The results of this study offer compelling evidence for the potential of AI-based voice analysis in the real-time assessment of psychological stress. Our findings demonstrate a statistically significant correlation between acoustic features extracted from speech and self-reported stress levels, suggesting that voice can serve as a reliable biomarker for stress detection. The AI models developed in this research, particularly those incorporating deep learning techniques, achieved promising accuracy in classifying different stress levels, indicating their effectiveness in capturing subtle vocal changes associated with psychological distress.

These findings have implications for various applications, including mental health monitoring, workplace stress management, and personalized interventions. The ability to passively and unobtrusively assess stress levels in real-time could enable early detection of psychological distress, facilitating timely intervention and preventing the escalation of mental health issues. Furthermore, the system's potential for integration into wearable devices and smartphone applications makes it a readily accessible tool for individuals seeking to manage their stress levels proactively.

However, it is crucial to acknowledge the limitations of the proposed system. The accuracy of the AI models is dependent on the quality and quantity of the training data. The current study utilized a specific dataset collected under controlled laboratory conditions, which may not fully represent the diversity of real-world scenarios and individual vocal characteristics. Further research is needed to evaluate the system's performance across different populations, languages, and environmental settings. Additionally, the system's reliance on acoustic features alone may not capture the full

complexity of psychological stress, which is influenced by a multitude of factors beyond vocal expression.

Compared to previous studies in the literature, our research builds upon existing knowledge by leveraging advanced AI techniques for real-time stress assessment. While previous studies have explored the relationship between voice and stress, many have relied on traditional machine learning algorithms or focused on offline analysis. Our study demonstrates the feasibility of using deep learning models to achieve higher accuracy and real-time performance, paving the way for more practical and scalable applications. For instance, studies using simpler statistical models achieved lower classification accuracies, typically around 60-70%, while our best performing model reached an accuracy of 85% on the test set. Furthermore, our system's ability to process speech in real-time distinguishes it from previous approaches that required lengthy processing times. Nevertheless, similar to other studies, our research acknowledges the need for further investigation into the generalizability and robustness of AI-based voice analysis for stress assessment. Future work should focus on addressing the limitations identified in this study and exploring the integration of other physiological and contextual data to enhance the accuracy and reliability of stress detection.

5.2. Advantages and Limitations

The AI-based voice analysis system for real-time stress assessment presents several key advantages. Objectivity is a primary benefit, as the system relies on quantifiable acoustic features extracted from speech, minimizing subjective biases inherent in traditional self-report measures or clinical observations. This objective assessment provides a more consistent and reliable evaluation of psychological stress levels. Immediacy is another significant advantage. The system offers real-time analysis, enabling immediate feedback and intervention opportunities. This contrasts sharply with methods requiring lengthy processing times or delayed results. Scalability is also a notable strength. Once trained, the AI model can be deployed across various platforms and applied to a large number of individuals simultaneously, making it suitable for widespread monitoring and stress management programs. The cost per assessment also decreases significantly as the system scales.

However, the system also faces several limitations. The performance of the AI model is heavily dependent on the availability of large and diverse datasets for training. Insufficient or biased training data can lead to inaccurate stress predictions, particularly for under-represented demographic groups. Sensitivity to noise is another concern. Background noise, variations in recording equipment, and individual differences in speech patterns can all affect the accuracy of voice feature extraction and subsequent stress assessment. Pre-processing techniques and robust feature selection are crucial to mitigate these effects, but complete elimination is challenging. Individual variability also presents a significant hurdle. Stress manifests differently in individuals, and the relationship between acoustic features and psychological stress can vary depending on factors such as personality, cultural background, and pre-existing medical conditions. Therefore, a universal model may not be equally accurate for all individuals, and personalized models or calibration techniques may be necessary to improve performance. The impact of factors like age, $gender$, and $culturalbackground$ on the model's accuracy requires further investigation. Sample Size limitation could add on: Given the relatively modest sample size, the results should be interpreted as indicative of feasibility and model potential rather than definitive performances benchmarks.

Add on: while classification accuracy provides a useful benchmark, stress is inherently continuous and subjective construct. Therefore, these results should be viewed as supporting relative discrimination rather than absolute diagnosis.

5.3. Potential Applications

The AI-based voice assessment system holds significant promise across diverse sectors. In mental health monitoring, the system could provide continuous, non-invasive stress level detection, alerting clinicians to potential crises or tracking the efficacy of therapeutic interventions. Imagine a scenario where subtle vocal changes, indicative of rising anxiety, trigger an automated message offering coping strategies or connecting the individual with support resources. Within workplace stress management, the system could identify high-stress periods or roles, enabling proactive adjustments to workload distribution or the implementation of targeted wellness programs. For example, if the average stress score, represented by S_{avg}, exceeds a threshold T, interventions could be automatically initiated. Emergency response scenarios could benefit from rapid stress assessment of first responders or victims, prioritizing aid and resource allocation based on real-time emotional state. Finally, personalized healthcare could leverage voice-based stress data to tailor treatment plans and lifestyle recommendations, enhancing patient adherence and overall well-being. The system's ability to provide objective, real-time stress measurements opens avenues for preventative care and improved outcomes in various domains.

6. Conclusion

6.1. Summary of Findings

This research investigated the feasibility and reliability of employing AI-based voice analysis for real-time psychological stress assessment. Our findings demonstrate a significant correlation between acoustic features extracted from speech and self-reported stress levels, as measured by standardized psychological scales. Specifically, we observed that features such as fundamental frequency (f_0), jitter, shimmer, and Mel-Frequency Cepstral Coefficients (MFCCs) exhibited statistically significant variations under induced stress conditions.

The AI models developed, including Support Vector Machines (SVM) and Recurrent Neural Networks (RNN), achieved promising accuracy in classifying stress levels based on voice data. The RNN model, leveraging temporal dependencies in speech, outperformed the SVM in capturing subtle stress-related vocal changes, achieving an average accuracy of 85% in distinguishing between stressed and non-stressed states. This suggests the potential of deep learning techniques for more nuanced stress detection.

A key contribution of this study lies in its real-time capability. The system was designed to process speech data in short segments, enabling near-instantaneous stress assessment. This real-time functionality opens avenues for proactive interventions and personalized stress management strategies. Furthermore, the non-invasive nature of voice analysis offers a convenient and unobtrusive alternative to traditional stress measurement methods, such as physiological sensors or questionnaires.

The significance of this research extends to various domains, including mental health monitoring, workplace stress management, and human-computer interaction. By providing a reliable and accessible means of assessing psychological stress, this technology can contribute to early detection of mental health issues and facilitate timely interventions, ultimately improving individual well-being and productivity. Future research will focus on refining the AI models, exploring the impact of individual differences on voice-based stress signatures, and validating the system's effectiveness in real-world settings.

6.2. Future Research Directions

Future research should prioritize enhancing the robustness and generalizability of AI models for voice-based stress assessment. Current models often exhibit performance degradation when applied to diverse populations, recording environments, and stressor types. Addressing this requires developing techniques that mitigate the impact of inter-individual variability in speech patterns and acoustic characteristics. Transfer learning

approaches, utilizing pre-trained models on large speech datasets, could be explored to improve generalization to new populations with limited labeled data.

Furthermore, investigating the incorporation of additional acoustic features beyond those traditionally used, such as non-linear dynamics and prosodic variations, may reveal subtle indicators of psychological stress. The exploration of advanced signal processing techniques, including wavelet transforms and time-frequency analysis, could uncover hidden patterns in the speech signal that are correlated with stress levels. Feature selection methods should be employed to identify the most relevant and informative acoustic features for stress detection.

Integrating contextual information, such as the speaker's background, current activity, and surrounding environment, can significantly improve the accuracy and interpretability of stress assessments. Contextual data, represented as C, can be incorporated into the model as additional input features, allowing the AI to better understand the speaker's emotional state. For example, knowing that a speaker is presenting at a conference ($C =$ conference presentation) can help differentiate between stress and excitement.

Large-scale validation studies are crucial to evaluate the real-world performance and clinical utility of these AI-based stress assessment systems. These studies should involve diverse populations and realistic scenarios to ensure that the systems are reliable and accurate in various contexts. The collection of physiological data, such as heart rate variability and cortisol levels, can provide objective measures of stress to validate the AI-based assessments.

Finally, ethical considerations and privacy implications must be carefully addressed. The development and deployment of AI-based stress assessment technologies should adhere to strict ethical guidelines and data privacy regulations. Anonymization techniques and secure data storage protocols should be implemented to protect the privacy of individuals. Transparency and explainability of the AI models are also essential to ensure that users understand how the assessments are made and to prevent potential biases. The potential for misuse of these technologies, such as in employment screening or surveillance, must be carefully considered and mitigated.

References

1. D. D. L. Veiga, T. M. Almeida, R. R. Uchida, and Q. Cordeiro, "The Fundamental Frequency of Voice as a Potential Stress Biomarker: A Systematic Review and Meta–Analysis," *Stress and Health*, vol. 41, no. 5, pp. e70112, 2025.
2. C. L. Giddens, K. W. Barron, J. Byrd-Craven, K. F. Clark, and A. S. Winter, "Vocal indices of stress: a review," *Journal of voice*, vol. 27, no. 3, pp. 390-e21, 2013.
3. G. A. Smith, "Voice analysis for the measurement of anxiety," *British Journal of Medical Psychology*, vol. 50, no. 4, pp. 367-373, 1977.
4. F. H. Fuller Jr, "Detection of emotional stress by voice analysis," No. LWLCR03B70, 1972.
5. R. Dillon, A. N. Teoh, and D. Dillon, "Voice analysis for stress detection and application in virtual reality to improve public speaking in real-time: A review," arXiv preprint arXiv:2208.01041, 2022.
6. K. R. Scherer, "Effect of stress on fundamental frequency of the voice," *The Journal of the Acoustical Society of America*, vol. 62, no. S1, pp. S25-S26, 1977.
7. H. J. Older, and L. L. Jenney, "Psychological stress measurement through voice output analysis," No. NASA-CR-141723, 1975.
8. L. J. Rothkrantz, P. Wiggers, J. W. A. Van Wees, and R. J. van Vark, "Voice stress analysis," in *International conference on text, speech and dialogue*, Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 449-456, 2004.
9. Arushi, R. Dillon, A. N. Teoh, and D. Dillon, "Detecting Public Speaking Stress via Real-Time Voice Analysis in Virtual Reality: A Review," in *Sustainability, Economics, Innovation, Globalisation and Organisational Psychology Conference*, Singapore: Springer Nature Singapore, pp. 117-152, 2023.
10. H. Hollien, "Vocal indicators of psychological stress," Forensic Psychology and Psychiatry/F. Wright, C. Bahn, RW Rieber (eds).—New York: New York Academy of Sciences, pp. 47-72, 1980.
11. S. Kanisha, E. N. Kumar, N. Charitha, B. Mehda, A. S. Vishnu, and R. Tilak, "Voice Based Stress Analysis and Detection Using Machine Learning," in *2024 10th International Conference on Advanced Computing and Communication Systems (ICACCS)*, vol. 1, IEEE, pp. 2231-2235, 2024.

12. R. Ruiz, C. Legros, and A. Guell, "Voice analysis to predict the psychological or physical state of a speaker," *Aviation, Space, and Environmental Medicine*, vol. 61, no. 3, pp. 266-271, 1990.