Detection of Mental Stress in Sports University Students through Machine Learning Techniques

Diptimoni Narzary[1*], Gypsy Nandi[1,2] , Uzzal Sharma[1,3] and Nongmaithem Suhinder Singh[1,4]
[1]Research Scholar, Department of Computer Applications, Assam Don Bosco University, Guwahati, India, diptinarzary56@gmail.com
[2]Associate Professor, Department of Computer Applications, Assam Don Bosco University, Guwahati, India
[3]Associate Professor, Department of Computer Science, Birangana Sati Sadhani Rajyik Viswavidyalaya, Golaghat, India
[4]Assistant Professor, Laxmibai National Institute of Physical Education NERC, Guwahati, Assam

Abstract

Mental stress is a significant issue affecting young individuals, particularly students, and can lead to serious cognitive disabilities, if left unaddressed. This study aimed to develop a machine-learning-based system to detect and measure stress levels in university-level sports students using vocal and acoustic features. Data were collected from 2400 students at the Lakshmibai National Institute of Physical Education (NERC), Guwahati, and analysed using Convolutional Neural Network (CNN) and random forest (RF) classification algorithms. The impact of exam pressure, match pressure, and recruitment stress on mental stress levels was examined. The performance of the algorithms was evaluated using the accuracy, precision, recall, and F1-score metrics. The RF algorithm achieved the highest accuracy (91.1 %) among the two classifiers. The proposed system aims to provide an objective tool for assessing stress levels, enabling earlier intervention, and more effective management of stress-related conditions by clinical psychologists. The study hypothesized that certain vocal characteristics, such as pitch variability, energy, Mel Frequency Cepstral Coefficients (MFCCs), Linear Predictive Coding Coefficients (LPCC), Zero Crossing Rate (ZCR), formant extraction, tempo beat extraction, and tonnetz extraction, would exhibit a significant correlation with higher stress levels. This study reviewed research on stress detection with machine learning, summarizing methods and classifier performance.

Keywords: Stress, Machine Learning, Sports, Mental Health, Acoustics.

1. Introduction

Stress is a common challenge faced by athletes, particularly sports students, who often experience intense physical and mental stress. Balancing academic responsibilities with rigorous training schedules can lead to increased stress levels, which if left unchecked, can negatively affect performance, health, and well-being. Research has shown that chronic stress among athletes can result in reduced focus, impaired decision making, and a higher risk of [1]. In the context of sports, the early detection and management of stress are crucial for ensuring optimal performance and maintaining mental and physical health. Traditional methods of stress detection, such as self-reporting and clinical evaluation, are often subjective and may not provide real-time insight. These methods rely on an individual's ability to recognize and articulate stress levels, which can be difficult, especially in high-performance sports environments [2] Consequently, there is a growing interest in leveraging technology, particularly machine learning, to develop objective real-time systems for stress detection. Machine learning techniques have shown significant promise in identifying stress through various physiological and behavioural signals, such as heart rate, skin conductance, and speech patterns [3]. In sports, vocal and acoustic signals have emerged as reliable sources for analysing stress because an athlete's vocal characteristics, including pitch, tone, and speech rate, can change under stress [4]. It is possible to detect early signs of stress using machine learning algorithms to analyse vocal features, thereby enabling timely intervention. This study sought to address the following research questions:

Q1: How accurately can machine learning models detect stress in sports students based on vocal characteristics?

Q2: Which vocal features are most indicative of stress in athletes?

Q3: How can the integration of a machine learning-based stress detection system improve stress management and performance in sports students?

Based on these research questions, this study tested the following hypotheses.

H1: Machine learning models can accurately detect stress in sports students by analysing vocal characteristics from audio recordings, thereby providing a reliable method for real-time stress detection.

H2: Specific vocal features, such as pitch variability, speech rate, and MFCC (Mel-Frequency Cepstral Coefficients), will show a significant correlation with elevated stress levels in athletes.

H3: Integrating a machine learning-based stress detection system into sports training will improve stress management, leading to enhanced performance and better mental health outcomes for sports students.

This study aimed to explore the application of machine learning for stress detection among sports students by focusing on vocal and acoustic signals. Given the unique stressors faced by this group, an accurate, real-time stress detection system could provide valuable insights for coaches, trainers, and mental health professionals, helping them design more effective training programs and mental health support systems.

2. Previous Work

Stress has long been recognized as a significant factor affecting athletic performance, particularly among student athletes, who must balance the competing demands of academics and high-level sports performance. Various studies have highlighted the importance of identifying and managing stress early to prevent adverse effects on both physical and mental health [2]. While traditional methods of stress assessment, such as self-report questionnaires and clinical evaluations, have been widely used, they often suffer from subjectivity and lack real-time applicability [5]. As a result, there has been growing interest in utilizing technology, specifically machine learning techniques, for more objective and efficient stress detection. A study by [6] examined the relationship between stress, coping strategies, and performance outcomes in sports and found that chronic stress can lead to burnout, reduced focus, and even increased injury risk. [7] further reported that unaddressed stress can cause fluctuations in motivation and recovery, ultimately hindering athletic progress. Given these findings, the ability to detect and manage stress in real time is critical for sports students, whose success depends on their ability to maintain optimal performance under pressure. While studying another paper related to sports, the author validated that the PSDQ-S, Exploratory Factor Analysis (EFA), and Confirmatory Factor Analysis (CFA) were crucial for confirming the nine-factor structure among 1,071 adolescents from Chihuahua, Mexico. The performance of these statistical algorithms was strong, with fit indices, such as GFI (.956), RMSEA (.041), and CFI (.975) indicating a good model fit. These results confirm the questionnaire's reliability and validity in assessing physical self-concept, demonstrating the effectiveness of the EFA and CFA in psychometric research [8].

Another study developed a model to predict sports practice behaviour among millennials based on their attitudes toward physical activity [9]. Using data from 1,141 individuals, the study identified a bi-dimensional structure of the attitude scale and effectively classified practicing and non-practicing individuals using a logistic regression. It was also found that millennials generally had a more positive attitude toward physical activity and sports than men. This study highlights the importance of fostering healthy habits among this generation.

In [10], the relationships between resilience, coping, commitment, and stress were examined among 190 male student-athletes aged 13-17. Using AI, they identified three clusters with similar psychological traits. Task-focused coping is linked to higher resilience and commitment, whereas emotion-focused coping increases stress levels. Student athletes tend to use the same coping strategies in both academic and sports environments, with differences based on their type, helping to identify at-risk groups for targeted interventions.

The study in [11] evaluated the impact of a gamified project on reducing anxiety about failure in Physical Education among 143 5th and 6th grade students. Using a mixed-methods approach, the results showed a significant decrease in post-intervention anxiety. The positive factors contributing to this reduction include passing tests, cooperative work, and the enjoyment of gamification. However, some challenges were noted, such as resistance work and stable groups, indicating areas for improvement of the gamified approach. Another study aimed to develop a scale to assess the sports values of Spanish middle and high school students related to sports. This study involved 1,316 students from 15 educational centers in Spain. Through exploratory and confirmatory factor analyses, the "Questionnaire of Sports and Olympic Values" was validated. This tool effectively measures the perception of sports-related values among students aged 12–18 [12].

In [13], the study examined life satisfaction (LS) among 723 Spanish Physical Education students aged 6–18 years. Using the Satisfaction with Life Scale (SLSS), the findings revealed similar LS levels across sexes. A negative correlation was observed between age and LS; however, BMI had no effect on LS. The location of the educational center slightly influenced LS, although the differences were not significant. Understanding these factors can help educators to promote healthy lifestyles and enhance adolescents' overall well-being.

In recent years, machine learning has emerged as a powerful tool for detecting stress through analysis of various physiological and behavioural data. [3] demonstrated the potential of machine learning models to accurately predict stress by analysing speech patterns, heart rate variability, and other bio signals. Their study found that models such as Support Vector Machines (SVM) and Random Forests (RF) achieved high accuracy in classifying stress levels when trained on appropriate datasets.

The studies cited in [4] and [14] explored the impact of vocal features on stress detection. They identified that vocal elements such as pitch, loudness, and Mel-Frequency Cepstral Coefficients (MFCCs) are dependable indicators of stress, capable of detecting emotional and physiological changes. These vocal biomarkers are especially useful in sports contexts, where swift, non-invasive stress detection techniques are crucial. Several research papers have detailed the creation of a stress detection system that utilizes heart rate variability (HRV) derived from ECG signals, incorporating analyses from both time and frequency domains. This system employs the k-nearest neighbour (KNN) algorithm for classification and has been evaluated on both Android devices and PCs. The authors observed that combining HRV features from both domains resulted in the most accurate stress representation, achieving a 79.17% accuracy rate with the KNN algorithm [15].

In a different study, researchers devised a cutting-edge technique for stress detection using EEG and ECG signals, termed Stress Recognition by Neuro analysis Safety and Risk Evaluation using Bayesian Networks (Se.Re.Ne.). The evaluation results demonstrated that the Se.Re.Ne. method, when combined with KNN, achieved a precision of 0.87, a recall of 0.71, and an f1-score of 0.78, leading to an overall accuracy of 68% [16]. Future research should focus on distinguishing between baseline (relaxed) and stress states by analyzing EEG sub band power ratios. This study employed both a Support Vector Machine (SVM) and k-nearest neighbour (KNN), with KNN showing the highest performance, reaching an accuracy of 99.42% [17].

Furthermore, another study proposed an advanced K-nearest neighbour machine learning model with improved feature selection to enhance the detection process. This model reduces input variables and eliminates noise through forward and backward filtering techniques. The enhanced KNN model predicted outcomes with 99% accuracy, establishing it as a highly effective classifier [18]. Similarly, the supervised machine-learning technique, Support Vector Machine (SVM), has been extensively and effectively utilized by numerous researchers. In a series of related studies, researchers have explored stress levels among university students and emotional recognition in speech using various machine learning algorithms. An analysis of Internet usage data from 206 JIIT Noida students found that SVM achieved the highest accuracy of 85.71% [19].

A research study focused on evaluating the effectiveness of a Support Vector Machine (SVM) classifier for emotion recognition in speech signals, comparing it with linear discriminant classifiers (LDC), k-nearest neighbour (KNN), and radial basis function neural network (RBFNN). The experiments, which utilized an emotional Chinese speech corpus, revealed that the SVM outperformed the other methods, achieving a peak accuracy of 85% [20]. Moreover, an SVM-based model designed for detecting speech emotions in children with autism spectrum disorder (ASD) achieved an accuracy of 77%. Researchers have also investigated the use of pitch variations and features such as MFCC, MEL, chroma, and Tonnetz, resulting in an 86.5% accuracy in emotion detection from speech [21]. Additional studies on Speech Emotion Recognition (SER) and mental stress detection using EEG signals demonstrated high performance, with SVM achieving accuracies of 96.36% and 94%, respectively [22]. An analysis of the influence of mental stress on students' Mental Stress-Handling Capability (MSHC) during online exams highlighted the SVM's effectiveness in predicting stress-handling capabilities [23]. Another study aimed at forecasting student employability initially achieved a 74.37% accuracy with SVM, which was improved by 13.43% through the application of customized hyperparameters influenced by the Teaching Learning Based Optimization (TLBO) algorithm [24].

3. Methodology

The methodology used in our research work for detecting and measuring stress levels in university-level sports students is discussed in detail in this section.

3.1 The data acquisition process

The data collection process was systematically structured to ensure the precision and dependability of the information obtained. The research focused on undergraduate students from the Lakshmibai

National Institute of Physical Education (LNIPE) NERC, Guwahati, with a sample size comprising 2400 students of both genders. The main aim was to measure the participants' stress levels using speech samples and the Beck's Depression Inventory (BDI), a widely acknowledged questionnaire for evaluating stress, anxiety, and depression. Initially, questions were crafted to assess the participants' emotional and psychological reactions to hypothetical situations they had faced in the previous month. These questions were designed to align with those in the BDI, ensuring consistency between self-reported data and vocal stress indicators.

During the interviews, participants responded to eight specific questions and provided speech samples while completing the BDI. Prior to the procedure, participants were thoroughly briefed on the study's purpose and assured of their privacy. Written informed consent was obtained from both the participants and the investigator, with both parties signing the consent form. Participants were given a copy of this form for their records. The collected data, including survey responses and BDI results, were assigned specific weights to compute stress level scores. After data collection, these results, along with the participants' speech samples, were submitted to clinical psychiatrist experts for further analysis and stress level prediction. Expert evaluation was essential in validating the machine learning models employed in the study, ensuring that predictions based on vocal indicators were consistent with the self-reported data.

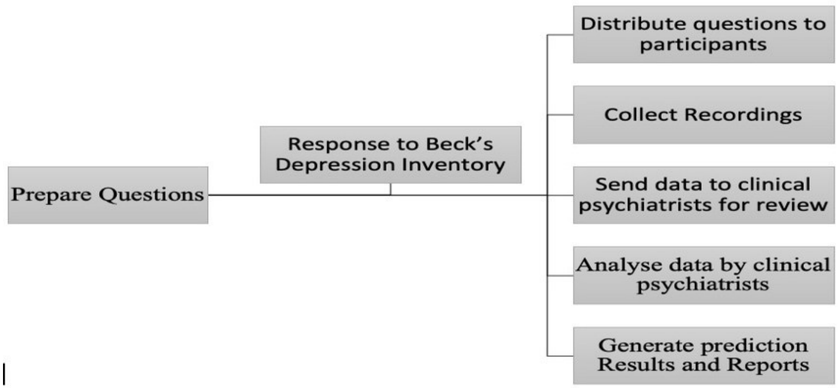


Fig1 : Workflow for detecting stress, incorporating the expertise of clinical psychiatrists experts

3.2 Machine Learning Approaches Used

The primary audio signals were converted into a feature set using standard speech processing techniques, including Mel-Frequency Cepstral Coefficients (MFCCs), Energy, Zero Crossing Rate (ZCR), Linear Predictive Coding Coefficients (LPCC), Pitch, Formant Extraction, Tempo Beat Extraction and Tonnetz Extraction. All features were normalized using min–max scaling to ensure that the input features were within a similar range, which improved the performance of the neural network. Below is an outline of how these features are integrated into machine learning approaches.

3.3 Preprocessing:

Speech signals were first pre-processed to normalize the amplitude and reduce background noise, ensuring consistency across recordings. This process enhances the clarity of the data by filtering out unwanted noise, allowing the features to better represent the true vocal characteristics.

3.4 Features Extraction:

There are key strategic elements that are paramount when performing analysis on audio signals, and one such element is feature extraction. It is the process of obtaining significant characteristics related to the analysis of the signal.

3.4.1 Mel-Frequency Cepstral Coefficients (MFCCs):

MFCCs are widely used in speech recognition tasks because of their ability to capture the shape of the speech spectra. In this process, a Fourier Transform was applied to convert the time-domain speech signal into the frequency domain. The Mel filter bank was then applied to the frequency spectrum, converting it into the Mel scale, which better reflects human auditory perception. Subsequently, the logarithm of the Mel-filtered signals was obtained and a Discrete Cosine Transform (DCT) was applied to obtain a compressed set of coefficients. These MFCCs represent the spectral properties of the speech signal, and are crucial for identifying stress-related variations [25].

The following equation combines the Fourier transform, Mel scale conversion, logarithmic compression, and DCT in a single expression to calculate the MFCCs.

$$\mathrm{C(n)} = \sum_{m=0}^{M-1} \left(log\left(\Sigma|X[k]|.\,H_m(k)\right)\right).cosx\left(\frac{\tau \mathrm{n}(2\mathrm{m}+1)}{2M}\right)$$

3.4.2 Energy:

Energy represents the loudness of the speech signal, and variations in the energy levels can indicate stress. Energy was computed by summing the squares of the speech signal over short frames. Stressful speech is often associated with energy irregularities, which may be reflected as fluctuations in loudness or intensity.

$$\mathrm{E} = \sum_{n=0}^{N-1} |x(n)|^2$$

Where N is the total number of samples in the frame, and x(n) represents the amplitude of the signal at time n.

3.4.3 Zero Crossing Rate (ZCR):

The ZCR measures the rate at which the speech signal changes its sign (crosses the zero line). This is a useful feature for detecting variations in speech patterns, such as those caused by emotional stress. ZCR was computed by counting the number of zero crossings within each frame of the speech signal. High ZCR values are often associated with noisy or turbulent speech and can indicate stress.

$$\mathrm{ZCR} = \frac{1}{N-1}\sum_{n=1}^{N-1} 1(x(n).\,x(n-1) < 0$$

3.4.4 Linear Predictive Coding Coefficients (LPCC):

LPCCs represent vocal tract configuration by modelling the speech signal using a linear predictive model. These coefficients were extracted by fitting a linear predictive model to the speech signal, in which the vocal tract shape was estimated from the speech waveform. LPCCs are particularly useful for capturing detailed information regarding vocal tract dynamics, which can vary significantly under stress [26].

$$\mathrm{x(n)} = \sum_{i=1}^{p} a_i . x(n-i) + e(n)$$

3.4.5 Pitch: It represents the perceived frequency of the sound, which can vary significantly under stress. Stress often creates pitch modulation or instability, and thus it serves as a crucial cue for analysis [27].

3.4.6 Formant Extraction: It captures the resonant frequencies of the vocal tract, aiding in the analysis of speech characteristics. Formant variation may mirror the change of vocal tract shape that results from either emotional or physical stress .

3.4.7 Tempo Beat Extraction: It measures the speed of the audio signal, which can influence the listener's emotional response. Stressful conditions can provoke irregular tempo patterns, reflecting agitation or nervousness [28].

3.4.8 Tonnetz Extraction: It represents harmonic relations in music, contributing to emotional interpretations of sound. Changes in Tonnetz features might indicate changes in emotion or stress-related tonal modification [29].

Table 1: Features considered for Mental Stress Detection

8 BASIC FEATURES	8 STATISTICAL VALUES FOR EACH FEATURES
MFCC	Mean
ENERGY	Standard deviation
LPCC	Median
ZCR	Minimum
PITCH	Maximum
FORMANT EXTRACTION	Range
TEMPO BEAT EXTRACTION	Skewness
TONNETZ EXTRACTION	Kurtosis

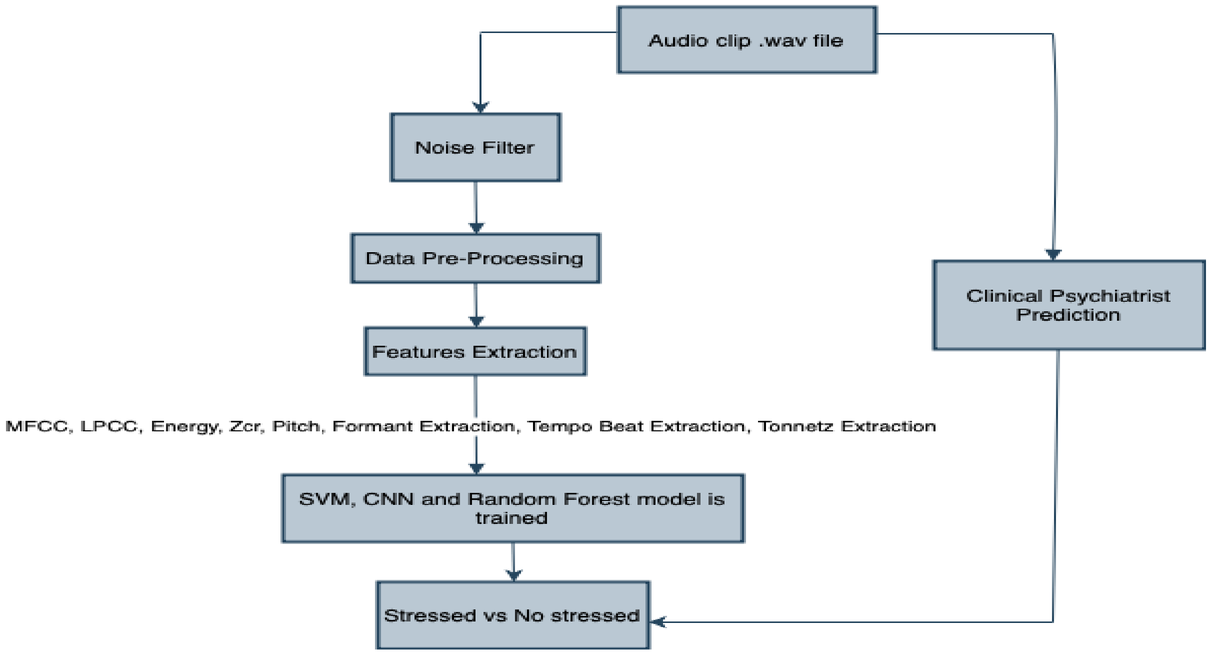


Fig 2: Various features extraction process

Speech samples for this study were collected from undergraduate students enrolled at the Lakshmibai National Institute of Physical Education (LNIPE) NERC, Guwahati. The sample included 2400 students, including both male and female participants. To evaluate stress levels, participants completed the Beck Depression Inventory (BDI), a globally recognized questionnaire. During the interviews, they were asked questions similar to those of the Beck Depression Inventory to obtain speech samples. Each participant responded to eight questions, expressed their perspectives. Prior to the study, participants were briefed about the purpose of the study and assured of their privacy. A consent form was signed by both the participant and the investigator, with a copy provided to the participant.

The questionnaire examined how participants responded to hypothetical scenarios they had faced in the past month. Their answers were given specific weights to determine their stress level scores. After data collection, voice recordings were sent to a clinical psychiatrist. An expert reviewed the recordings and questionnaires for each participant to assess whether they were experiencing stress. The expert assessments were then compared with the results generated by the proposed model algorithm to validate the accuracy and reliability of the system.

3.5 Experimental Evaluations

This section describes the environmental setup and the associated results and comparative analysis of the experiments carried out with the proposed approach.

3.6 Environmental Setup

In this study, two distinct model architectures were employed for stress detection: a Convolutional Neural Network (CNN) and Random Forest classifier. The CNN was designed to leverage spatial hierarchies within audio feature representations, facilitating the learning of complex patterns that are indicative of stress. Its architecture comprises multiple convolutional layers, including an initial layer with 64 filters and a subsequent layer with 128 filters, responsible for feature extraction. To enhance the training process, batch normalization was applied after each convolutional layer to stabilize the outputs and promote convergence. Max pooling layers reduced the spatial dimensions of feature maps, while a flatten layer prepared the pooled outputs for fully connected layers. The model incorporated two dense layers with 128 and 64 neurons, respectively, alongside dropout layers to mitigate overfitting by randomly deactivating the neurons during training. The final output layer utilizes a sigmoid activation function to predict the likelihood of the input being classified as stressed. Trained over 100 epochs with a batch size of 32, the CNN achieved an accuracy of 81.6%, precision of 77.7%, recall of 73.9%, and an F1 score of 75.7%. Simultaneously, a Random Forest classifier was implemented as a benchmark for comparison. This ensemble learning method combines multiple decision trees to enhance the prediction accuracy and robustness. The dataset was split into training (80%) and testing (20%) sets, and the model was trained using 100 estimators. The Random Forest classifier demonstrated impressive performance metrics, with an accuracy of 91.1%, precision of 91.7%, recall of 84.7%, and F1 score of 88.1%.

Model 1: Convolutional Neural Network (CNN)

The CNN architecture was designed with the following layers:

- **Convolutional Layers**: Two Conv1D layers with ReLU activation, followed by Batch Normalization and Max Pooling layers to extract features from the input audio signals.
- **Dense Layers**: Fully connected layers with dropout for regularization and L2 regularization to prevent overfitting.
- **Output Layer**: A single neuron with a sigmoid activation function for binary classification [30].

i. Convolution Operation:

$$Z_{ij}^{(k)} = \sum_{m=1}^{F}\sum_{n=1}^{C} X_{(i+m)(j+n)}.\, W_{mn}^{(k)} + b^{(k)}$$

Where, $Z_{ij}^{(k)}$: Output of the convolutional layer for the k[th] filter.

$X_{(i+m)(j+n)}$: Input features.

$W_{mn}^{(k)}$: Filter weights.

$b^{(k)}$: Bias term.

ii. Batch Normalization:

$$BN(Z) = \gamma.\frac{Z-\mu}{\sqrt{\sigma^2+\epsilon}} + \beta$$

Where, Z: Input to the batch normalization layer.

μ: Mean of the batch.

σ^2: Variance of the batch.

γ, β: Learnable scale and shift parameters.

ϵ: Small constant to avoid division by zero.

iii. Pooling: Max-pooling reduces dimensionality:

$$P_{ij} = \max Z_{(i+m)(j+n)}$$

Where, P_{ij}: Pooled output.

$Z_{(i+m)(j+n)}$: Input values in the pooling window.

iv. Fully Connected Layers:

$$Y = f(W.X + b)$$

Where, W: Weight matrix.
X: Flattened input from previous layers.
b: Bias Term.
f: Activation function

Training and Evaluation

The model was trained for 100 epochs with a batch size of 32. The Adam optimizer was used, with a learning rate of 0.0001. The performance of the model was evaluated using accuracy, precision, recall, and F1-score metrics. The Fig 3, confusion matrix of CNN showed that the model effectively distinguished between "Stressed" and "Not Stressed" classes.

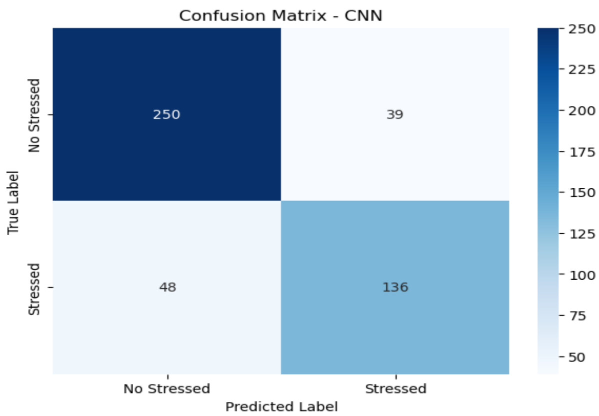
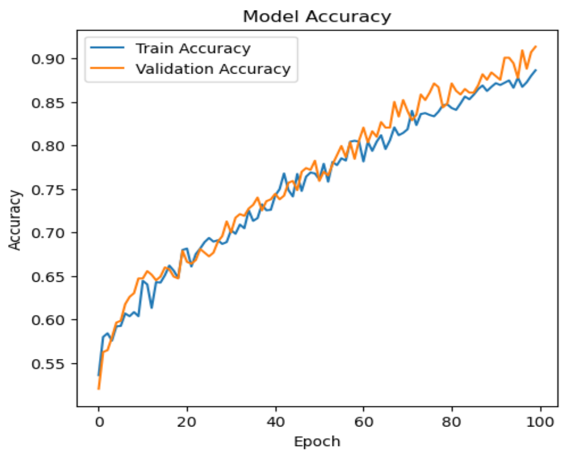


Fig 3: The confusion matrix of CNN

Figures 4 and 5 present two key graphs illustrating the training and validation performances of the CNN model over 100 epochs. Both figures track the performance of the model during the training and validation phases, providing insight into how well the model is learning.



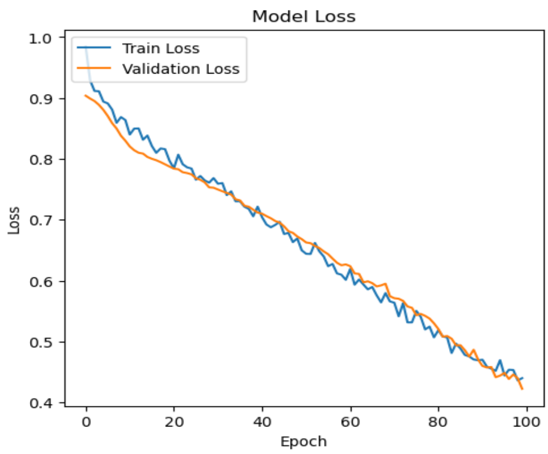


Fig 4: The training and validation accuracy over the epochs. Fig 5: The training and validation loss over the epochs.

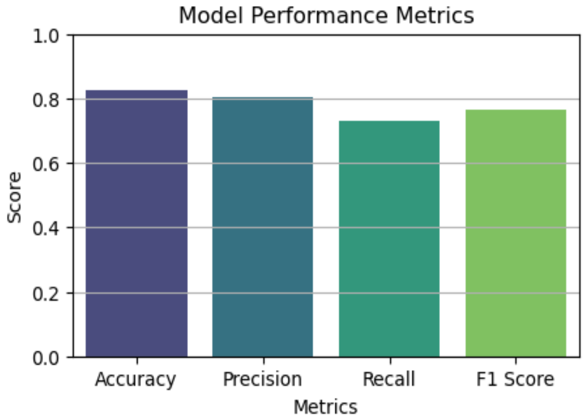


Fig 6: Model Performance Metrics of CNN

The values for these metrics are presented as bars, showing the performance of the model across all key evaluation criteria. The y-axis represents the score (ranging from 0 to 1) and the x-axis lists the different metrics.

Model 2: Random Forest Classifier
Model Implementation

A Random Forest classifier is an ensemble method that combines multiple decision trees to make predictions. Each tree generates a prediction for the input data and the final output is determined by majority voting among all trees [31]. This reduced overfitting and increased the robustness of the model. In this implementation, 100 decision trees are used, and the Random Forest algorithm aggregates their predictions to classify the data as "Stressed" or "Not Stressed."
The following equation was used in the forest to predict the output:

$$h_i(X) = Tree_i(X)$$

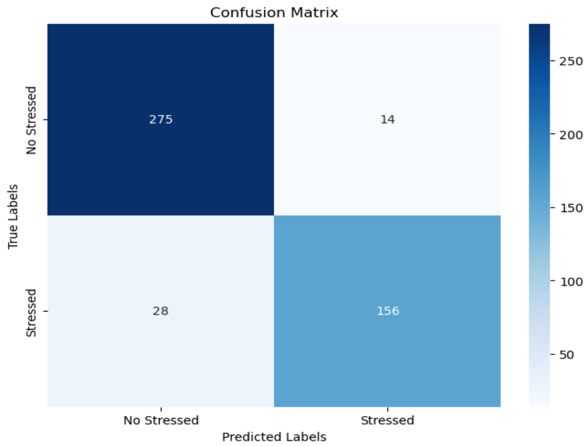
Where, $h_i(X)$: Prediction of the i^{th} tree.
X: input features.

The Random Forest aggregates predictions using majority voting:

$$H(X) = mode\{h_1(X), h_2(X), \dots, h_N(X)\}$$

Where, H(X): Final prediction.
mode: Most common prediction among N trees.

Training and Evaluation
Similar to the CNN, the dataset was split into training and testing sets. The model was trained on the same features and evaluated using the same metrics.



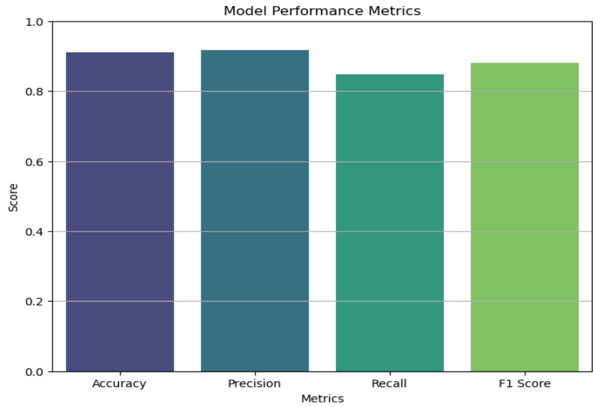


Fig 7: The Confusion matrix of RF

Fig 8: Model performance metrics of RF

Fig. 7 presents a confusion matrix used to evaluate the performance of a classification model by comparing the true labels ("Not Stressed" and "Stressed") with the predicted labels. The matrix shows that 275 cases were correctly predicted as "Not Stressed" and 156 were correctly identified as "Stressed." However, there were 14 false positives, where "Not Stressed" was incorrectly predicted as "Stressed," and 28 false negatives, where "Stressed" was predicted as "Not Stressed."
Fig. 8 displays the key performance metrics of the model: Accuracy, Precision, Recall, and F1 Score. All metrics had scores close to 0.85, indicating that the model achieved a high level of accuracy and balanced performance.

4. Experimental Results and Discussion

The main goal of this study was to determine student stress levels. Performance of stress detection models using key metrics, comparative analyses, and simulations. The models employed were a Convolutional Neural Network (CNN) and Random Forest classifier, both assessed using metrics such as accuracy, precision, recall, and F1 score. The CNN achieved an accuracy of 81.6%, precision of 77.7%, recall of 73.9%, and F1 score of 75.7%. These results indicate that the CNN effectively identified the stress patterns, although some misclassifications occurred, particularly in borderline cases, as reflected by the slightly lower recall. The ability to capture complex audio features contributes to its excellent performance.

In contrast, the Random Forest classifier demonstrated superior performance, achieving an accuracy of 91.1%, precision of 91.7%, recall of 84.7%, and an F1 score of 88.1%. This model's high precision suggests minimal false positives, whereas its strong recall indicates accurate identification of stressed individuals. The balance between precision and recall, as indicated by the F1 score, highlights the robustness and reliability of the RF in stress detection tasks. A comparative analysis of the confusion matrices revealed that while the CNN was effective in distinguishing between stressed and non-stressed samples, it exhibited a higher rate of false positives and negatives than the Random Forest, which achieved a clearer separation between the two classes.

Simulations conducted on the CNN showed steady convergence of the training and validation accuracy over 100 epochs, with minor overfitting controlled through dropout and batch normalization. The training and validation loss curves stabilized after approximately 80 epochs, indicating effective learning. For the Random Forest classifier, simulations were performed with multiple random seeds to ensure robustness, demonstrating a consistently high performance across various data partitions. A feature importance analysis revealed that MFCCs, energy, Zero Crossing Rate (ZCR), pitch, format extraction, tempo beat extraction, and tonnetz extraction were the most influential factors in predicting the stress levels.

Table 2: Results of the testing

Classifiers	Convolutional Neural Network (CNN)	Random Forests (RF)
Accuracy	81.6%	91.1%
Precision	77.7%	91.7%
Recall	73.9%	84.7%
F1-score	75.7%	88.1%

5. Conclusion and Future Scope

This study focuses on detecting stress using speech samples and evaluating the performance of two machine learning models: a Convolutional Neural Network (CNN) and a Random Forest classifier. Speech data were collected from 2,400 undergraduate students at the Lakshmibai National Institute of Physical Education (LNIPE), Guwahati, alongside stress assessment using Beck's Depression Inventory (BDI). The audio features extracted, such as MFCCs, Energy, Zero Crossing Rate (ZCR), LPCC, Pitch, Formant Extraction, Tempo Beat, and Tonnetz, were processed and used to train and evaluate both models.

The CNN leveraged its ability to capture complex patterns in audio features and achieved reasonable performance with an accuracy of 81.6%, precision of 77.7%, recall of 73.9%, and F1 score of 75.7%. However, it exhibited limitations in handling certain misclassifications, as revealed by confusion matrix analysis. In comparison, the Random Forest classifier consistently outperformed the CNN across all metrics, achieving an accuracy of 91.1%, precision of 91.7%, recall of 84.7%, and an F1 score of 88.1%. Its ensemble learning approach allows for better generalization, resulting in fewer false positives and false negatives.

The confusion matrix analysis further highlighted the superiority of Random Forest in correctly classifying stressed and non-stressed individuals, with higher true-positive and true-negative rates. Although both models demonstrated the potential for stress detection using speech data, the Random Forest classifier emerged as a more reliable and accurate model. Future research could explore hybrid approaches, combining the CNN's feature extraction capabilities with the Random's decision-making power, to further enhance performance. Additionally, expanding the dataset and incorporating more diverse speech samples could improve the generalizability and robustness of the model in real-world scenarios.

The findings of this study have significant practical implications for sports psychology and the mental health of university students. This study provides a valuable tool for the early identification of mental stress among athletes by developing a machine-learning-based system that utilizes vocal and acoustic features to detect stress levels [32]. This early detection is crucial as it allows for timely intervention, which can prevent the progression of stress into more severe psychological conditions, such as anxiety or depression, that could negatively impact both academic and athletic performance.

6. Conflict of Interest

In recent years, scientists have been actively pursuing research in this area. The inclusion of expert clinicians and psychosocial professionals in the current project is expected to lead to enhanced results upon its successful conclusion. The ongoing investigation shows potential for patent acquisition. This marks a notable advancement in the field, providing multiple benefits across various areas of study. As a result, there is a strong likelihood of securing a patent in the coming years.

References

[1] K. Smith, "Suffering in Sport," 2019, pp. 121–140. doi: 10.1108/S1476-285420190000012010.

[2] A. Postema, H. van Mierlo, A. B. Bakker, and M. T. (Mariska) Barendse, "Study-to-sports spillover among competitive athletes: a field study," *Int J Sport Exerc Psychol*, pp. 1–24, Apr. 2022, doi: 10.1080/1612197X.2022.2058054.

[3] Z. Fan *et al.*, "Environmental enrichment modulates HPA axis reprogramming in adult male rats exposed to early adolescent stress," *Neurosci Res*, vol. 172, pp. 63–72, Nov. 2021, doi: 10.1016/j.neures.2021.04.007.

[4] X.-D. Yang *et al.*, "Stress during a Critical Postnatal Period Induces Region-Specific Structural Abnormalities and Dysfunction of the Prefrontal Cortex via CRF1," *Neuropsychopharmacology*, vol. 40, no. 5, pp. 1203–1215, Apr. 2015, doi: 10.1038/npp.2014.304.

[5] Z. Zainudin, S. Hasan, S. M. Shamsuddin, and S. Argawal, "Stress Detection using Machine Learning and Deep Learning," *J Phys Conf Ser*, vol. 1997, no. 1, p. 012019, Aug. 2021, doi: 10.1088/1742-6596/1997/1/012019.

[6] M. Kaiseler, J. M. Poolton, S. H. Backhouse, and N. Stanger, "The Relationship Between Mindfulness and Life Stress in Student-Athletes: The Mediating Role of Coping Effectiveness and Decision Rumination," *Sport Psychol*, vol. 31, no. 3, pp. 288–298, Sep. 2017, doi: 10.1123/tsp.2016-0083.

[7] K. E. Smith and S. D. Pollak, "Early life stress and development: potential mechanisms for adverse outcomes," *J Neurodev Disord*, vol. 12, no. 1, p. 34, Dec. 2020, doi: 10.1186/s11689-020-09337-y.

[8] M. Ornelas Contreras, M. Conchas Ramírez, A. C. Díaz-Leal, L. H. Blanco Ornelas, and Y. S. Rangel Ledezma, "Propiedades psicométricas de una versión informatizada del Physical Self-Description Questionnaire en adolescentes mexicanos (Psychometric properties of a computerized version of the Physical Self-Description Questionnaire in Mexican adolescents)," *Retos*, no. 37, pp. 22–26, Jul. 2019, doi: 10.47197/retos.v37i37.68900.

[9] A. Mujika-Alberdi, J. J. Gibaja-Martins, and I. García-Arrizabalaga, "Un modelo predictivo del comportamiento de la práctica deportiva entre la generación Millennial (A predictive model of sports practice behavior among the Millennial generation)," *Retos*, no. 39, pp. 129–136, Jun. 2020, doi: 10.47197/retos.v0i39.77903.

[10] H. C. Bellod, V. Buendía Ramón, E. Carballeira Fernández, and J. F. Guzmán Luján, "Analysis of stress and academic-sports commitment through Self-organizing Artificial Neural Networks (Análisis del estrés y el compromiso académico-deportivo mediante Redes Neuronales Artificiales Auto-organizativas)," *Retos*, vol. 42, pp. 136–144, Oct. 2021, doi: 10.47197/retos.v42i0.86983.

[11] B. Rodríguez Martín, G. Flores Aguilar, and J. Fernández Río, "Ansiedad ante el fracaso en educación física ¿puede la gamificación promover cambios en las alumnas de primaria? (Anxiety about failure in physical education. Can gamification promote changes in elementary school girls?)," *Retos*, vol. 44, pp. 739–748, Apr. 2022, doi: 10.47197/retos.v43i0.90864.

[12] X. Pena-Pérez, I. Portelas-Pino, J. González-Silva, J. P. Fuentes-García, and M. J. Martínez-Patiño, "Diseño y validación de un cuestionario en español de valores deportivos y olímpicos (Design and validation of a questionnaire in Spanish on sports and Olympic values)," *Retos*, vol. 47, pp. 408–418, Nov. 2022, doi: 10.47197/retos.v47.93726.

[13] C. Galán-Arroyo, S. Gómez-Paniagua, A. Castillo-Paredes, and J. Rojo-Ramos, "Satisfacción Vital y Adolescencia en la Educación Física Española (Life Satisfaction and Adolescence in Spanish Physical Education)," *Retos*, vol. 55, pp. 65–71, Mar. 2024, doi: 10.47197/retos.v55.103739.

[14] X. Yang *et al.*, "Perceived psychological stress and associated factors in the early stages of the coronavirus disease 2019 (COVID-19) epidemic: Evidence from the general Chinese population," *PLoS One*, vol. 15, no. 12, p. e0243605, Dec. 2020, doi: 10.1371/journal.pone.0243605.

[15] P. D. Purnamasari, R. Martmis, and R. R. Wijaya, "Stress Detection Application based on Heart Rate Variability (HRV) and K-Nearest Neighbor (KNN)," in *2019 International Conference on Electrical Engineering and Computer Science (ICECOS)*, IEEE, Oct. 2019, pp. 271–276. doi: 10.1109/ICECOS47637.2019.8984436.

[16] D. Virmani, A. Minocha, L. Goyal, M. Malhotra, and M. Gupta, "Se.Re.Ne.: Stress Detection Using EEG and ECG," 2021, pp. 189–198. doi: 10.1007/978-981-15-9712-1_16.

[17] T. H. Priya, P. Mahalakshmi, V. Naidu, and M. Srinivas, "Stress detection from EEG using power ratio," in *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*, IEEE, Feb. 2020, pp. 1–6. doi: 10.1109/ic-ETITE47903.2020.401.

[18] A. S. Maria, R. Sunder, and A. M. Antony, "Stress Prediction Using Enhanced Feature Selection and KNN Model," in *2023 Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA)*, IEEE, Jan. 2023, pp. 1–5. doi: 10.1109/ACCTHPA57160.2023.10083348.

[19] R. Ahuja and A. Banga, "Mental Stress Detection in University Students using Machine Learning Algorithms," *Procedia Comput Sci*, vol. 152, pp. 349–353, 2019, doi: 10.1016/j.procs.2019.05.007.

[20] T. Seehapoch and S. Wongthanavasu, "Speech emotion recognition using Support Vector Machines," in *2013 5th International Conference on Knowledge and Smart Technology (KST)*, IEEE, Jan. 2013, pp. 86–91. doi: 10.1109/KST.2013.6512793.

[21] R. Matin and D. Valles, "A Speech Emotion Recognition Solution-based on Support Vector Machine for Children with Autism Spectrum Disorder to Help Identify Human Emotions," in *2020 Intermountain Engineering, Technology and Computing (IETC)*, IEEE, Oct. 2020, pp. 1–6. doi: 10.1109/IETC47856.2020.9249147.

[22] K. V. Krishna, N. Sainath, and A. M. Posonia, "Speech Emotion Recognition using Machine Learning," in *2022 6th International Conference on Computing Methodologies and Communication (ICCMC)*, IEEE, Mar. 2022, pp. 1014–1018. doi: 10.1109/ICCMC53470.2022.9753976.

[23] R. Gupta, M. A. Alam, and P. Agarwal, “Modified Support Vector Machine for Detecting Stress Level Using EEG Signals,” *Comput Intell Neurosci*, vol. 2020, pp. 1–14, Aug. 2020, doi: 10.1155/2020/8860841.

[24] S. Jayachandran and B. Joshi, “Customized support vector machine for predicting the employability of students pursuing engineering,” *International Journal of Information Technology*, vol. 16, no. 5, pp. 3193–3204, Jun. 2024, doi: 10.1007/s41870-024-01818-w.

[25] U. Sharma, “3rd INTERNATIONAL CONFERENCE ON INNOVATIVE COMPUTING AND COMMUNICATION (ICICC-2020) Detection o f D epression f rom S peech Si gnal through L inear SV M us ing MFCC Feature.” [Online]. Available: https://ssrn.com/abstract=3562977

[26] P. Kumar, “Automatic Speaker Recognition using LPCC and MFCC,” *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 3, no. 4, pp. 2106–2109, 2015, doi: 10.17762/ijritcc2321-8169.150474.

[27] A. Koduru, H. B. Valiveti, and A. K. Budati, “Feature extraction algorithms to improve the speech emotion recognition rate,” *Int J Speech Technol*, vol. 23, no. 1, pp. 45–55, Mar. 2020, doi: 10.1007/s10772-020-09672-4.

[28] S. Dixon, “Automatic Extraction of Tempo and Beat From Expressive Performances,” *J New Music Res*, vol. 30, no. 1, pp. 39–58, Mar. 2001, doi: 10.1076/jnmr.30.1.39.7119.

[29] E. J. Humphrey, T. Cho, and J. P. Bello, “Learning a robust Tonnetz-space transform for automatic chord recognition,” in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, Mar. 2012, pp. 453–456. doi: 10.1109/ICASSP.2012.6287914.

[30] R. K and V. R. Murthy Oruganti, “Stress Detection using CNN Fusion,” in *TENCON 2021 - 2021 IEEE Region 10 Conference (TENCON)*, IEEE, Dec. 2021, pp. 492–497. doi: 10.1109/TENCON54134.2021.9707438.

[31] V. Rajendra Mali and Prof. A. R. Surve, “STRESS PREDICTION USING RANDOM FOREST LINEAR REGRESSION,” *International Journal of Engineering Applied Sciences and Technology*, vol. 6, no. 3, Jul. 2021, doi: 10.33564/IJEAST.2021.v06i03.048.

[32] J. C. Mundt, P. J. Snyder, M. S. Cannizzaro, K. Chappie, and D. S. Geralts, “Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology,” *J Neurolinguistics*, vol. 20, no. 1, pp. 50–64, Jan. 2007, doi: 10.1016/j.jneuroling.2006.04.001.