International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 10s, 2025
https://theaspd.com/index.php

A Machine Learning-Based Approach For Stress Detection In Sports Students Using Vocal Analysis

Narzary Diptimoni[1*]
[1]Research Scholar, Department of Computer Applications, Assam Don Bosco University, Guwahati, 781017, India
diptinarzary56@gmail.com; https://orcid.org/0009-0005-7632-0339

Nandi Gypsy[2]
Department of Computer Applications, Assam Don Bosco University, Guwahati, 781017, India
gypsy.nandi@dbuniversity.ac.in; https://orcid.org/0000-0001-9417-4439

Sharma Uzzal[3]
Department of Computer Science, Birangana Sati Sadhani Rajyik Viswavidyalaya, Golaghat, 785621, India
druzzalsharma@gmail.com; https://orcid.org/0000-0002-5264-1016

Basumatary Shankar Jyoti[4]
Dean & Associate Professor, LNIPE, NERC Tepesia Guwahati, 782402, Assam, India
sbasumatary27@gmail.com

Abstract
Sports students frequently encounter considerable stress due to the dual demands of intense physical training and academic responsibilities. Conventional methods for detecting stress, such as self-reports and clinical evaluations, tend to be subjective and unsuitable for real-time applications. This study investigates the use of machine learning to objectively detect stress among different students from sports background examining their vocal traits. Speech samples were gathered from 670 undergraduates at the North Eastern Regional Centre (NERC) of the Lakshmibai National Institute of Physical Education (LNIPE), which is located in Sonapur, near Guwahati, and stress levels were confirmed using the Beck Depression Inventory (BDI). Essential acoustic features, such as Mel-Frequency Cepstral Coefficients (MFCC), pitch, Zero-Crossing Rate (ZCR), and Formant frequencies, were derived from the audio recordings. Three machine learning algorithms - Support Vector Machine (SVM), K-Nearest Neighbours (KNN), and Long Short-Term Memory (LSTM) RNN - were utilized to categorize stress levels. The findings showed that KNN surpassed the other models, achieving the highest accuracy (94.72%) and F1-score (94.67%), followed by LSTM at 90.90% accuracy. The SVM exhibited the lowest performance (62.01% accuracy), underscoring its challenges in managing intricate vocal stress features. These results indicate that machine learning-based vocal analysis offers a promising method for real-time stress detection in sports students, potentially facilitating early intervention and better stress management strategies.
Keywords: *Sports students, Stress detection, Machine learning, Acoustic feature, KNN, SVM, LSTM*

INTRODUCTION

Stress is a significant challenge faced by athletes, particularly sports students, who must balance intense physical training with academic responsibilities (Saleh et al., 2017). The demands of rigorous training schedules, competition, and academic pressures can lead to heightened stress levels, which, if left unmanaged, can negatively impact performance, well-being, and overall mental health. Research has shown that chronic stress among athletes can impair cognitive functions, such as focus, decision-making, and reaction time, ultimately affecting athletic performance and increasing the risk of injury (Kumar & Ankayarkanni, 2022). Thus, early detection and management of stress in sports students are crucial for ensuring optimal performance and long-term well-being.

Traditional stress detection methods, such as self-report questionnaires and clinical evaluations, are often subjective and rely heavily on an individual's ability to recognize and articulate their stress levels (Narzary

et al., 2025). These approaches may not provide real-time insights, making them less effective for athletes who require immediate and data-driven stress assessments to adjust their training routines accordingly. 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 demonstrated significant promise in detecting stress through various physiological and behavioural signals such as heart rate, skin conductance, and speech patterns (Shanbhog M & Medikonda, 2023). Among these, vocal and acoustic signals have emerged as reliable indicators of stress because an individual's vocal characteristics, including pitch, tone, and speech rate, tend to change under stress. Analysing these vocal features using machine learning algorithms can enable early stress detection, allowing for timely intervention and improved stress management strategies in athletes.
This study aimed to explore the application of machine learning to stress detection in sports students by analysing vocal and acoustic signals. Specifically, the research was guided by the following key questions.

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

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

Q3: In what ways can incorporating a machine learning driven stress detection system enhance stress management and boost performance among sports students?

To address these 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 Mel-Frequency Cepstral Coefficients (MFCC), 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 employed multiple machine-learning models, including Support Vector Machines (SVM), LSTM Recurrent Neural Networks (RNN), and K-Nearest Neighbours (KNN), to classify stress levels based on speech signals. The model utilized key audio features, including MFCC, Pitch, Zero-Crossing Rate (ZCR), Formant Extraction, Tempo Beat Extraction, and Tonnetz Extraction to assess stress in athletes. Additionally, techniques such as data augmentation and the synthetic minority oversampling technique (SMOTE) are applied to improve the robustness of the model and address class imbalances in the dataset. The dataset used in this study was pre-processed by applying noise filtering, feature extraction, and standardization techniques before being fed into the models for training and evaluation.
The model put forward seeks to deliver a scalable, real-time approach for identifying stress in sports students, assisting coaches, trainers, and mental health experts in creating more effective training and mental health support systems. By bridging the gap between traditional stress assessment methods this study contributes to the advancement of machine learning applications in sports psychology and athletic well-being. These findings could lead to the development of an automated stress monitoring system that enables early detection, timely intervention, and personalized stress management strategies, ultimately enhancing both the mental resilience and athletic performance of sports students.

Background

In recent years, there has been a growing interest in detecting stress through the use of machine learning and deep learning methods. Numerous classifiers, such as K-Nearest Neighbours (KNN), LSTM Recurrent Neural Networks (RNN), and Support Vector Machines (SVM) have been widely utilized to evaluate stress levels by examining physiological, textual, and audio features. This review explores previous studies on stress detection and analyses the performance of various classification models. Machine learning has been extensively used for stress classification across various modalities, including physiological signals, speech, and textual data.
Enrique Garcia Ceja, Venet Osmani, and Oscar Mayora examined smartphone-based stress detection using accelerometer data. They conducted an 8-week study with 30 subjects from two different organizations, where stress levels were self-reported three times daily. Using statistical models, they achieved a maximum accuracy of 71% for user-specific models and 60% for similar-user models, relying

solely on accelerometer data [(Garcia-Ceja et al., 2016). This research underscores the promise of using smartphone sensors for detecting stress, while also highlighting the necessity for enhanced data modelling methods.

Similarly, in 2020 Bobade & Vani proposed some deep learning techniques for stress detection using multiple datasets from wearable sensors (Bobade & Vani, 2020). Their machine learning approaches achieved accuracy rates of up to 81.65% for classifying three categories and 93.20% for binary classification. Meanwhile, their deep learning models reached accuracies of 84.32% and 95.21%, respectively.

Another study focused on speech analysis, such as that by Gupta et al. (Gupta Megha V & Vaikole Subhangi, n.d.) (n.d.), demonstrated the effectiveness of CNNs for stress detection, achieving accuracies between 94.26% and 94.3%. Wearable sensor-based stress detection has been a growing research area. Schmidt et al. (Schmidt et al., 2018) introduced a dataset named WESAD, that integrates multiple modals for the purpose of stress and affect detection, where DT, KNN, RF, LDA, SVM, and AdaBoost achieved accuracies of 80% and 93%. Sandulescu et al. (2015) also employed SVM on wearable physiological sensors and achieved an accuracy of 82% (Sandulescu et al., 2015). Ahuja and Banga (2019) focused on stress detection using ECG, respiration, skin conductance, and EMG signals, with KNN achieving 80% accuracy (Ahuja & Banga, 2019).

Emotional recognition has also been explored for stress detection. Pandey (2017) used physiological responses to music to classify emotional arousal, achieving 70% accuracy (Pandey, 2017). Ferdinando et al. (2016) applied SVM for emotion recognition based on heart rate variability, but the accuracy was relatively low at 48.5% (Ferdinando et al., 2016). By comparing different machine learning models, Ghaderi et al. (2016) found that KNN and SVM provided high accuracies of 92.06% and 96.82%, respectively, for stress detection using physiological signals (Ghaderi et al., 2016).

Mumtaz et al. (2018) evaluated multiple classifiers, including MLP, DT, KNN, SVM, and deep learning, and reported that decision trees had the best performance, with 95% accuracy (Mumtaz et al., 2018; Rosales et al., 2019). Similarly, Rosales et al. (2019) compared linear regression, Naïve Bayes, RF, and SVM for stress detection among university students, where the highest accuracy achieved by SVM around 85.71% (Rosales et al., 2019).

Text-based stress detection has also been explored. (Lin et al., 2017) Lin et al. (2017) developed a stress detection system using natural language processing and machine learning techniques, where SVM achieved an accuracy of 90%, with a recall of 90%, precision of 94%, and an F1-score of 92%. Hossain et al. (2022) studied stress detection in social media users using SVM, Naïve Bayes, Decision Trees, and Random Forest, with SVM provided the highest accuracy of 75% and an F1-score of 80% (Hossain et al., 2022). Furthermore, Nijhawan et al. (2022) proposed a physiological-based smart stress detector using LR, KNN, and SVM, where SVM achieved the highest accuracy of 95 to 96.67% (Nijhawan et al., 2022).

Additionally, Swaymprabha Alias Megha Mane and Arundhati Shinde proposed a novel architecture, Stress Net, that integrates a 2D CNN with a LSTM network for stress detection in EEG signals. These EEG signals were broken down into their alpha, beta, and theta components, which were subsequently transformed into images of azimuthal projection. Azimuthal projection-based images were processed using a 2D CNN for feature extraction, followed by an LSTM to capture the temporal dependencies. Using fully connected layers the model classifies the different states of stress. Evaluated on the SEED and DEEP datasets, Stress Net outperformed the traditional human stress detection methods, achieving an impressive accuracy of 97.8% (Mane & Shinde, 2023).

Phutela et al. (2023) proposed a stress classification system using EEG signals acquired from a 4-electrode Muse EEG headband. This study analysed EEG signals from thirty-five volunteers who were exposed to stress- and non-stress-inducing movie clips. The stress classification model compared Multilayer Perceptron (MLP) and Long Short-Term Memory (LSTM) networks. The highest accuracy of 93.17% was achieved using a two-layer LSTM architecture, thereby demonstrating the effectiveness of deep learning in EEG-based stress detection(Phutela et al., 2022).

Acikmese and Alptekin (2023) explored passive mobile phone sensor data for stress detection using deep learning approaches. They employed LSTM, CNN, and CNN-LSTM models on the Student Life dataset,

which contained passive mobile sensing data and stress feedback from college students. Their LSTM model achieved an accuracy of 62.83% for 460 test instances, although the study highlighted limitations owing to the small dataset size. The findings suggest that while LSTM models show promise, larger and more diverse datasets are required to improve model generalization (Acikmese & Alptekin, 2019).
Another study, Md. Saif Hassan Onim and Hima explored the integration of behavioral time series sensor data, including Electrodermal Activity (EDA), Blood Volume Pulse (BVP), and Skin Temperature (ST), for stress prediction. Using RNNs and LSTM networks, the relationship between cortisol biomarker levels and physiological signals was examined. Their study, conducted on 40 older adults, reported an impressive accuracy of 94% for RNN and 96% for LSTM models, highlighting the effectiveness of time-series-based machine learning in stress monitoring applications (Onim & Thapliyal, 2024).

METHODOLOGY

This research utilized a machine-learning technique to identify stress in sports students by analysing vocal features from audio recordings. The process involves several phases, such as gathering data, preprocessing, extracting features, choosing a model, training and evaluating it, and analysing its performance. Figure 1 illustrates the schematic flow diagram of the proposed stress detection system, highlighting the sequential process from audio data acquisition to feature extraction, model training, and stress classification.

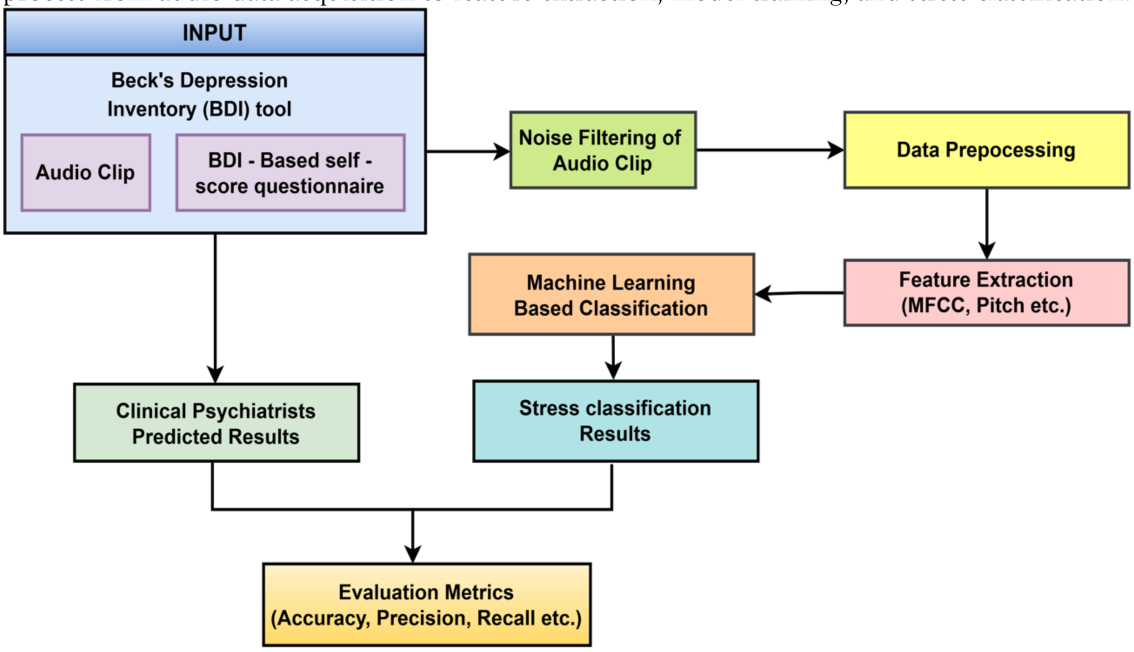

Fig 1: Schematic Flow Diagram of Stress Detection

Data Collection

For this research the speech samples were collected from UG students currently enrolled at the North Eastern Regional Centre (NERC) of the Lakshmibai National Institute of Physical Education (LNIPE), which is located in Sonapur, near Guwahati, Assam. The sample dataset consists of 670 number of both male and female students. Participants were administered the Beck Depression Inventory (BDI), an internationally acknowledged questionnaire, to evaluate their stress levels. The collected speech samples and responses to the questionnaire were cross-matched to validate the predictions of stress levels based on the speech analysis.
To gather their speech samples during the interview process, some particular questions were asked which are similar to those in the BDI. The participants answered eight questions, expressing their perspectives. Before the procedure, participants were briefed about the study and guaranteed confidentiality. Both the participant and the investigator signed a consent form, with a copy given to the participant for their records.Participants were questioned about their responses to hypothetical scenarios that they had faced in the previous month during the surveys. Their responses were assigned specific weights to generate the

stress level scores. For the interview recordings, a set of predefined questions were asked in a calm environment, and the voice samples were stored for analysis. The main goal of cross-matching these datasets is to verify and validate the stress-level predictions made from speech samples.The dataset for this study comprises audio recordings of sports students under normal (non-stressed) and stressed states. The stress conditions were induced through controlled experiments that involved intense training sessions, cognitive stressors, such as academic problem-solving tasks, and self-reported stress assessments. A clinical psychiatrist validated the stress labels that were assigned to each audio sample. The collected data were stored in ***.wav*** format, ensuring high-quality audio recordings for analysis. Each recording session included multiple speakers and the dataset was balanced to include variations in speech patterns, pitch, and tone across different individuals.

Data Preprocessing

To ensure the quality and usability of the dataset, several pre-processing steps were applied.

Noise Filtering: Digital noise reduction methods, including spectral subtraction and adaptive filtering, are used to eliminate the background noise and other unwanted disturbances.

Data Augmentation: To increase dataset diversity and robustness, augmentation techniques, such as adding Gaussian noise and time stretching, were applied.

Handling of Missing Values: Any missing or erroneous data entries were identified and removed.

Standardization: The audio recordings were normalized to ensure uniformity in the feature extraction.

3.3 Feature Extraction

Using signal processing techniques, the vocal and acoustics features were extracted from the pre-processed audio recordings. The audio dataset contained key extracted features as shown in Table 1:

Table 1: Key Speech Features Used in Stress Detection

Features	**Descriptions**
MFCC (Mel-Frequency Cepstral Coefficients)	Represents the short-term power spectrum of sound, capturing the phonetic characteristics of speech.
Energy	Refers to the overall power or amplitude of the audio signal, which can indicate loudness or intensity.
LPCC (Linear Predictive Cepstral Coefficients)	Used to model the vocal tract and analyse speech characteristics.
ZCR (Zero-Crossing Rate)	The rate at which the audio signal changes sign (crosses zero). It is often used to differentiate between voiced and unvoiced speech.
Pitch	Relates to the perceived frequency of sound, which is important in determining the tone or pitch of speech.
Formant extraction	Formants are resonant frequencies in speech that give clues about vowel sounds and speaker characteristics.
Tempo beat extraction	Extracts the tempo or rhythm of the sound, which could indicate stress based on speaking speed.
Tonnetz extraction	Captures harmonic and tonal relationships in the sound, often used in music analysis but can provide additional information about speech patterns.

These features were extracted using digital signal processing techniques and stored numerically for further machine learning analysis. After extracting the features, a sample of the extracted feature results is shown in Figure 2.

FinalDataset2

SL No	No. of Speaker	MFCC	ENERGY	ZCR	LPCC	PITCH	Formant_Extraction	Tempo_beat_extraction	Tonnetz_extraction	CLASS
1	Speaker 1	-36.14	0.00266103	0.129912	-4.28	0.023219955	-0.04609856	-0.037547699	0.022009662	No Stressed
2	Speaker 2	-24.83	0.003432253	0.071045	-4.86	0.034829932	0.001579352	0.026721633	0.021424521	No Stressed
3	Speaker 3	-39.98	0.002432898	0.132276	-5.06	0.046439909	-0.019909547	0.04496526	-0.051819857	Stressed
4	Speaker 4	-25.10	0.00361177	0.159505	-6.22	0.058049887	0.005900421	-0.028479466	-0.042446431	Stressed
5	Speaker 5	-18.56	0.004282814	0.078868	-3.82	0.069659864	0.002360712	-0.045383505	0.017354478	No Stressed
6	Speaker 6	-24.40	0.004123062	0.083435	-3.16	0.081269841	0.004263064	0.073911854	0.009774297	Stressed
7	Speaker 7	-33.24	0.003500273	0.122932	-2.31	0.092879819	0.00944114	0.02877317	0.002000608	Stressed
8	Speaker 8	-30.67	0.003500575	0.096031	-8.35	0.104489796	-0.004419052	-0.029971981	0.070785112	Stressed
9	Speaker 9	-24.34	0.003490938	0.091668	-3.03	0.116099773	0.006194677	0.066711334	0.032117995	Stressed
10	Speaker 10	-33.16	0.00318017	0.026758	-3.74	0.127709751	0.005478572	0.032158024	0.023737594	Stressed
11	Speaker 11	-32.15	0.003252164	0.071615	-4.21	0.139319728	0.004417275	-0.001421166	-0.006464445	No Stressed
12	Speaker 12	-35.12	0.002556429	0.070117	-2.89	0.150929705	-0.011741108	-0.052152839	-0.047616271	Stressed
13	Speaker 13	-30.29	0.003695102	0.045329	-2.6	0.162539683	0.000631348	-0.010218891	0.01337657	Stressed
14	Speaker 14	-29.00	0.002949332	0.109253	-4.45	0.17414966	0.015087075	0.071004471	0.044193321	No Stressed
15	Speaker 15	-28.35	0.003570956	0.067464	-2.72	0.185759637	-0.004322345	-0.011266558	0.045799493	No Stressed
16	Speaker 16	-27.94	0.003201027	0.05558	-4.52	0.197369615	-0.011240681	-0.0821146	0.013272137	Stressed
17	Speaker 17	-24.82	0.003812739	0.074423	-6.88	0.208979592	0.001164538	0.01052316	0.037338575	No Stressed
18	Speaker 18	-21.80	0.003838922	0.094251	-4.85	0.220589569	-0.00214096	0.05077285	-0.044709065	No Stressed
19	Speaker 19	-38.44	0.002873107	0.088293	-4.51	0.232199546	0.014249894	-0.022323271	0.040862103	Stressed
20	Speaker 20	-27.36	0.003557663	0.09196	-1.12	0.243809524	0.006170757	0.007820308	-0.026384249	No Stressed
21	Speaker 21	-28.63	0.003071604	0.053467	-3.45	0.255419501	0.010467169	-0.039134651	-0.080200476	No Stressed
22	Speaker 22	-28.63	0.00411088	0.095749	-4.36	0.267029478	0.01347131	-0.061656621	-0.016395662	No Stressed
23	Speaker 23	-31.35	0.002963066	0.033936	-3.98	0.278639456	-0.000573587	-0.002917996	-0.021085572	No Stressed
24	Speaker 24	-32.39	0.002861087	0.107572	-4.44	0.290249433	0.002431764	0.018834332	0.025107948	Stressed
25	Speaker 25	-31.15	0.003359705	0.085348	-6.54	0.30185941	0.013446709	-0.031193588	-0.032045788	Stressed
26	Speaker 26	-32.13	0.003949312	0.018945	-4.14	0.313469388	-0.006504805	-0.009773414	0.02182637	No Stressed
27	Speaker 27	-19.66	0.004209109	0.110107	-5.02	0.325079365	-0.001691667	0.033653681	-0.012894695	Stressed
28	Speaker 28	-22.68	0.003984225	0.100868	-3.86	0.336689342	-0.01047385	-0.02199048	0.025286555	No Stressed
29	Speaker 29	-29.32	0.004066918	0.073892	-2.32	0.34829932	-0.028007004	-0.005198326	0.03591405	Stressed
30	Speaker 30	-21.63	0.003911378	0.144007	-9.07	0.359909297	0.046576103	0.096946309	-0.044513509	Stressed

Fig 2: Sample of Extracted Audio Features Used for Stress Detection.

Experimentation And Simulations

The experimentation phase involved designing and implementing a machine learning pipeline for stress detection in sports students using vocal characteristics. The entire process was executed in a controlled experimental environment to ensure reliable and repeatable results. In this study, three classification algorithms were used to detect stress in sports students based on speech features. Each model follows a distinct mathematical approach to learn patterns from the extracted features and classify speech samples into stressed or non-stressed categories. Below are the classification algorithms along with their mathematical formulations.

Support Vector Machine (SVM)

SVM is a supervised learning algorithm that finds the optimal hyperplane to separate different classes in a high-dimensional space. It aims to maximize the margin between the closest points (support vectors) of different classes.

For the classification, the following formula is used:

$$< ww.\, xx > + bb0 \geq 1, \forall yy = 1 \quad (1)$$

$$< ww.\, xx > + bb0 \leq -1, \forall yy = -1 \quad (2)$$

Where (x, y) represents a pair of training sets and b0 stands for the bias condition (Narzary & Sharma, n.d.).

The trained SVM model was assessed using standard performance metrics, including accuracy, precision, recall, and F1-score. Figure 3 illustrates the confusion matrix for the SVM model used in stress detection, showcasing its classification performance. The model successfully identified 166 non-stressed individuals (true negatives) but mistakenly classified 213 non-stressed individuals as stressed (false positives). Additionally, it correctly recognized 304 stressed individuals (true positives) while incorrectly labelling 75 stressed cases as non-stressed (false negatives). This distribution of predictions highlights the model's strengths in correctly identifying stressed individuals while struggling with distinguishing non-stressed cases, which could affect its overall precision and recall.

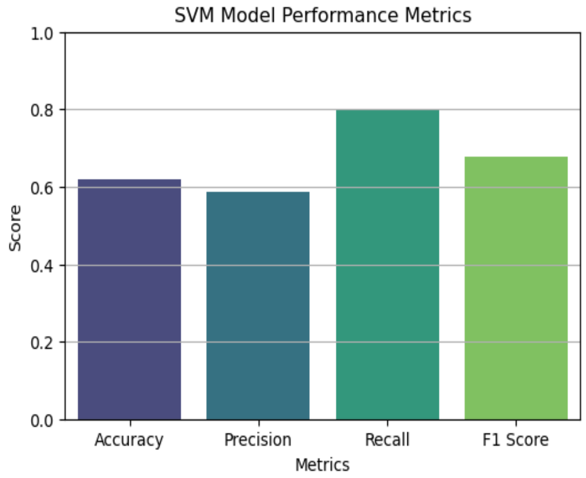

Fig 3: Performance Metrics of SVM Model for Stress Detection.

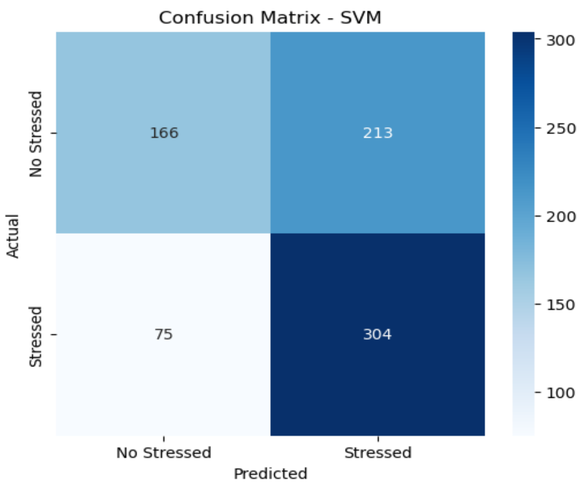

Fig 4: Confusion Matrix of SVM Model for Stress Detection.

K-Nearest Neighbours (KNN)

In this study, K-Nearest Neighbors (KNN) was implemented as one of the machine learning models to classify speech samples into stressed and non-stressed categories based on extracted vocal features.
The decision for classification is determined by a majority vote among the K nearest training samples.
The distance between a test sample x and a training sample x_i was calculated using the Euclidean distance, which is given by:

$$\boldsymbol{d(x, x_i)} = \sqrt{\sum_{j=1}^{n}(x_j - x_{ij})^2} \quad (3)$$

where x_j represents the j[th] feature of the test sample, $\boldsymbol{x_{ij}}$ represents the j-th feature of the training sample, and n is the total number of extracted features (Tatiur Rahman, 2015). After computing distances, the algorithm selects the K nearest samples, and the final classification label $\boldsymbol{\hat{y}}$ is determined through majority voting, expressed as:

$$\hat{y} = \arg\max \sum_{i=1}^{K} I(y_i = y) \quad (4)$$

Where $I(y_i = y)$ is an indicator function that counts the occurrences of class y among the K nearest neighbors.
To enhance KNN's classification performance, hyperparameter tuning was conducted using GridSearchCV. The optimal K value was determined by testing different values such as 3, 5, 7, 9, 11, 13, and 15, with the best K selected based on accuracy and F1-score. The distance metric used was Euclidean distance, as it yielded superior results compared to Manhattan distance for this dataset. Additionally, a uniform weighting scheme was applied, ensuring that all K neighbors contributed equally to the classification decision.
Below is the performance metrics and confusion matrix for the KNN model, visually representing its classification results:

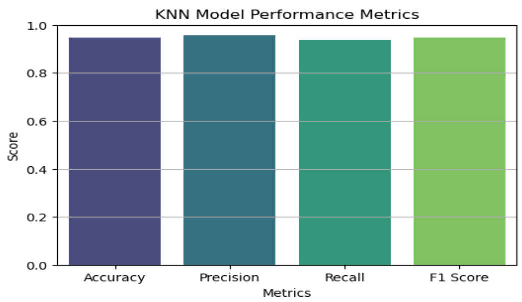

Fig 5: KNN Model Performance Metrics for Stress Detection.

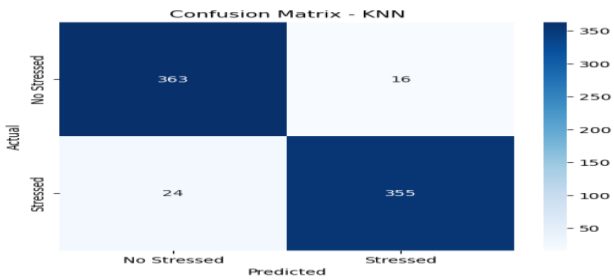

Fig 6: Confusion Matrix for KNN Model in Stress Detection.

Long Short-Term Memory (LSTM)

In this study, Long Short-Term Memory (LSTM) networks were employed to classify speech samples into stressed and non-stressed categories by leveraging the temporal dependencies present in vocal features. Unlike traditional classifiers such as KNN and SVM, which treat each audio feature independently, LSTM effectively models the sequential patterns in speech by maintaining long-range dependencies, making it particularly suitable for analysing time-series data like speech signals (Winata et al., 2018).

LSTM is a type of Recurrent Neural Network (RNN) that overcomes the vanishing gradient problem by incorporating memory cells that regulate information flow through input, forget, and output gates. The mathematical formulation of LSTM is as follows:

Forget Gate: Determines which information should be discarded from the previous cell state:

$$\boldsymbol{f_t = \sigma(W_f.[h_{t-1}, x_t] + b_f)} \quad (5)$$

Where f_t is the forget gate activation, W_f and $b_f)$ are learnable weights and biases, h_{t-1} is the hidden state from the previous step, and x_t is the current input.

Input Gate: Decides which new information should be stored in the cell state:

$$\boldsymbol{i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i)} \quad (6)$$

$$\boldsymbol{C_t = tanh(W_C.[h_{t-1}, x_t] + b_C)} \quad (7)$$

where i_t is the input gate activation, and C_t represents the candidate cell state.

Cell State Update: It combines previous and new information:

$$\boldsymbol{C_t = f_t.C_{t-1} + i_t.C_t} \quad (8)$$

Where C_t is the updated cell state.

Output Gate: Determines the hidden state h_t which influences the next time step:

$$\boldsymbol{o_t = \sigma(W_o.[h_{t-1}, x_t] + b_o)} \quad (9)$$

$$\boldsymbol{h_t = o_t}.\tanh(\boldsymbol{C_t}) \quad (10)$$

Where o_t is the output gate activation and h_t is the new hidden state (Mane & Shinde, 2023).

The LSTM network was trained using the cross-entropy loss function, which is well-suited for classification tasks, and the Adam optimizer, known for its adaptive learning rate and efficient convergence. To enhance model performance, hyperparameter tuning was conducted using GridSearchCV, systematically exploring multiple configurations to identify the optimal values. The number of LSTM units was varied among 64, 128, and 256, with 128 units providing the best balance between model complexity and performance. Similarly, different batch sizes of 16, 32, and 64 were tested, and a batch size of 32 yielded the most stable and efficient training. The learning rate, a critical parameter controlling weight updates, was tuned across 0.001, 0.005, and 0.01, with 0.001 proving optimal for minimizing loss and improving generalization. Additionally, dropout regularization was applied to prevent overfitting, with dropout rates of 0.2, 0.3, and 0.5 tested; a dropout rate of 0.3 was found to be the most effective in balancing regularization while maintaining model performance. These optimized hyperparameters ensured that the LSTM network effectively captured temporal dependencies in speech features, making it a robust classifier for stress detection.

A batch size of 32 and 150 epochs are selected based on validation performance. The model employs the binary cross-entropy loss function because it is designed to handle binary classification tasks, such as distinguishing between stressed and not stressed.

The training process involves monitoring accuracy and loss across both training and validation datasets. The training accuracy and loss graphs demonstrate that the model achieves convergence effectively, as evidenced by the simultaneous improvement in both validation and training accuracy, which suggests minimal overfitting.

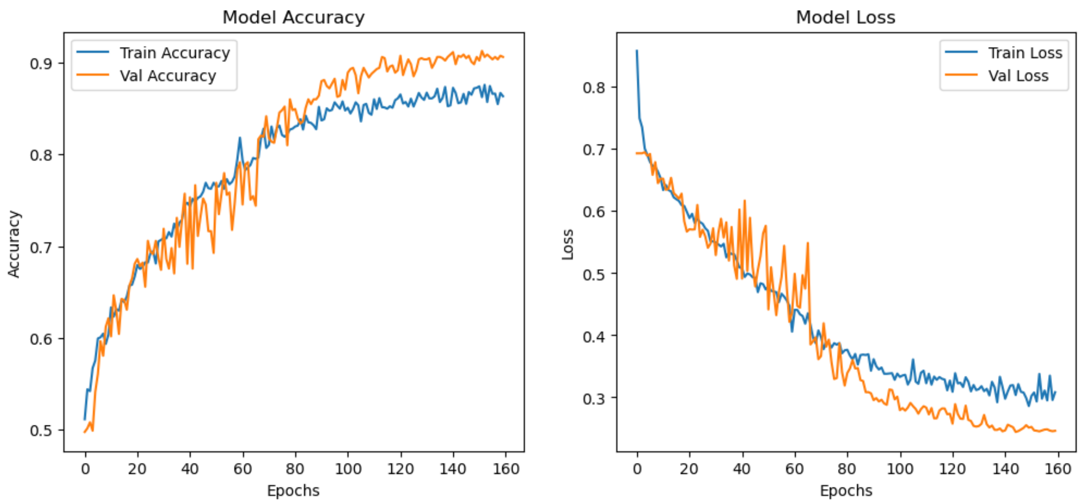

Fig 7: Training and Validation Accuracy & Loss Curves for LSTM Model.

The confusion matrix shows that the model correctly classifies most instances, with 336 true negatives and 353 true positives and 43 false positives and 26 false negatives. Additionally, the bar chart summarizing model performance metrics confirms that accuracy, precision, recall, and F1-score all remain consistently high, supporting the model's effectiveness this study.

The model is evaluated using standard classification metrics, including accuracy, precision, recall, and F1-score. The confusion matrix provides insights into misclassification rates. The results are summarized as follows:

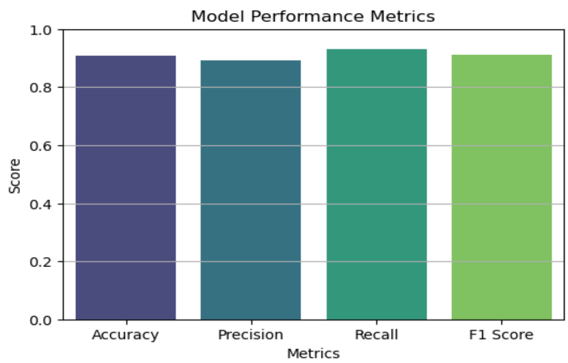

Fig 8: LSTM Model Performance Metrics for Stress Detection.

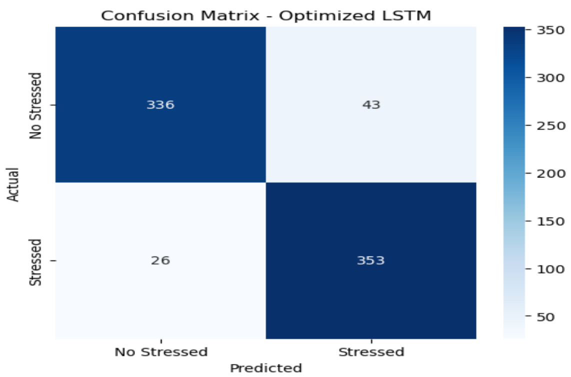

Fig 9: Confusion Matrix for LSTM Model in Stress Detection.

Data Augmentation and Preprocessing

Data augmentation and preprocessing are crucial steps in ensuring that the machine learning models generalize well to real-world data. In this study, two augmentation techniques were applied to enhance the dataset, followed by preprocessing steps to standardize and normalize the extracted features. Since the dataset consisted of vocal recordings under different stress conditions, data augmentation techniques were used to artificially expand the dataset, improving model robustness and reducing overfitting (Mumuni & Mumuni, 2022). The following augmentation methods were applied:

Gaussian Noise Addition: To make the model more robust to background noise variations, random Gaussian noise was added to the speech features (EL Bilali et al., 2021). The noisy signal was generated using the formula:

$$\boldsymbol{x_{augmented} = x + N(0, \sigma^2)} \qquad (11)$$

Where x is the original feature vector, and $N(0, \sigma^2)$ represents Gaussian noise with zero mean and variance σ^2.

Feature Scaling: Random scaling factors were applied to speech features to simulate variations in vocal intensity and microphone sensitivity (Ozsahin et al., 2022). This was performed using:

$$\boldsymbol{x_{augmented} = x \times \propto} \qquad (12)$$

Where $\propto$ is a random scaling factor in the range [0.9,1.1], ensuring small variations in amplitude without distorting the speech signal.

These augmentation techniques were applied systematically to generate multiple variations of each original sample, significantly increasing the dataset size and improving the model's ability to generalize.

Results and Discussions:

The effectiveness of three machine learning models Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and LSTM Recurrent Neural Network (RNN) was assessed using four critical performance metrics: Accuracy, Precision, Recall, and F1-Score. These metrics provide a comprehensive evaluation of each model's ability to classify stress levels accurately.

The summarized results are presented in Table 2.

Table 2: Classification Results of SVM, KNN, and LSTM (RNN) Models

Classifier	Accuracy	Precision	Recall	F1-Score
SVM	62.01%	58.80%	80.21%	67.86%
KNN	94.72%	95.69%	93.67%	94.67%
LSTM (RNN)	90.90%	89.14%	93.14%	91.10%

Performance Comparison

The comparative analysis of the classifiers reveals that the KNN model significantly outperformed both SVM and LSTM (RNN) across all performance metrics. KNN achieved the highest accuracy of 94.72%, indicating its robustness in distinguishing stress levels in vocal data. Additionally, its high precision of 95.69% suggests that it had minimal false positives, making it highly reliable for stress classification. Furthermore, with a recall of 93.67% and an F1-score of 94.67%, KNN demonstrated a strong balance between identifying true stress instances and avoiding misclassification.

The LSTM (RNN) model followed closely behind with an accuracy of 90.90%, which is still significantly higher than SVM. Notably, LSTM exhibited a recall of 93.14%, the highest among the three classifiers. This suggests that LSTM was highly effective in capturing and correctly identifying stress instances from the vocal features. However, its precision of 89.14%was slightly lower than KNN's, meaning it had a slightly higher rate of false positives. Nevertheless, its F1-score of 91.10% confirms its ability to maintain a good trade-off between precision and recall.

Conversely, the SVM model delivered the weakest performance, achieving an accuracy of 62.01%, which is considerably lower than both KNN and LSTM. Although its recall of 80.21% was relatively high, the precision of 58.80% suggests that the model struggled with false positives, resulting in an F1-score of 67.86%. This suggests that SVM had difficulty generalizing well for stress detection from vocal data and was not as effective in distinguishing stress and non-stress instances compared to KNN and LSTM.

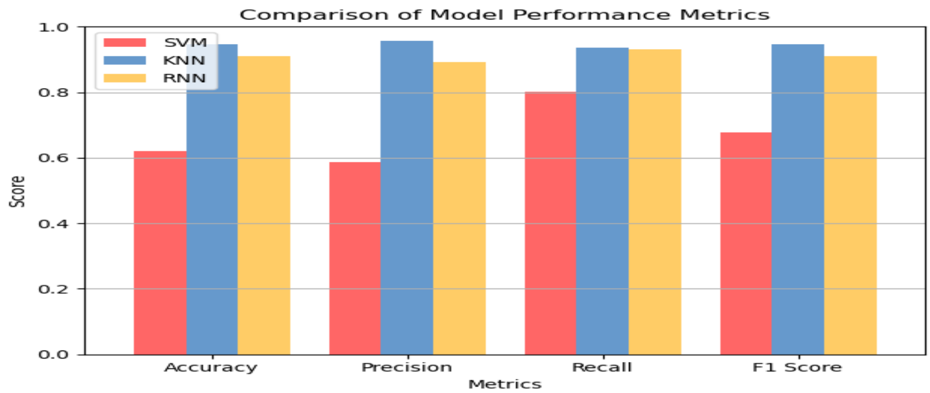

Fig 10: Comparison of Classification Results Across Models.

Implications of Findings

The strong performance of KNN suggests that stress-related vocal features exhibit clear separability in the feature space, making KNN's distance-based classification approach highly effective. This model's success could be attributed to the nature of vocal stress patterns, which may form distinct clusters in high-dimensional space, making nearest neighbour based classification particularly advantageous.

The effectiveness of LSTM further highlights the importance of sequence learning in vocal analysis. Since LSTMs are designed to capture temporal dependencies and patterns, their ability to achieve a high recall suggests that stress signals in voice recordings may have sequential characteristics that the model successfully identified. This finding supports the notion that deep learning techniques can enhance stress detection by leveraging time-dependent patterns in speech.

On the other hand, the underwhelming performance of SVM indicates that a linear or kernel-based approach may not be optimal for this dataset. The model's relatively low precision and overall accuracy suggest that stress-related vocal features are not linearly separable, which limits SVM's effectiveness. This highlights the need for advanced feature engineering or the use of non-linear classifiers when dealing with complex and high-dimensional datasets like vocal stress signals.

CONCLUSION

The findings from this study indicate that the KNN classifier is the most effective model for stress detection based on vocal features, achieving the highest accuracy and F1-score. The LSTM model also performed well, particularly in recall, which suggests its potential for applications where correctly identifying stressed individuals is crucial. However, SVM underperformed, highlighting its limitations in handling vocal stress features without further optimization.

These results contribute to ongoing research on machine learning applications for stress detection, particularly among sports students. Future research should explore larger datasets, ensemble learning approaches, and advanced feature extraction techniques to further enhance the accuracy and reliability of the stress classification models. The integration of multimodal data sources, such as physiological signals combined with vocal features, may also improve model performance in real-world scenarios.

Conflict of Interest

In recent years, in this are researchers have pursued advancements. By involving both clinical and psychosocial experts in the ongoing project, it is anticipated that the research will yield improved outcomes once it is successfully completed. Ongoing research has indicated the possibility of obtaining a patent. This represents a significant advancement in the field, providing numerous advantages across different research areas. As a result, it is highly probable that a patent will be obtained in the future.

Acknowledgement

I wish to convey my deepest appreciation to everyone who offered support and guidance throughout this research. First, I am profoundly thankful to my supervisor for their invaluable guidance, encouragement, and unwavering support during the study. Their insightful feedback and constructive suggestions have significantly improved the quality of this work. I am also thankful to the faculty and staff of LNIPE NERC, Guwahati, for providing the essential resources and assistance needed during the data-collection phase. Special thanks go to the students who participated in the study and contributed their speech samples for the analysis. I would also like to acknowledge the contributions of clinical and psychosocial experts, whose expertise greatly enriched the research design and interpretation of the results. Finally, I extend my gratitude to my family and friends for their steadfast support and motivation without which this work would not have been possible.

REFERENCES

1. Acikmese, Y., & Alptekin, S. E. (2019). Prediction of stress levels with LSTM and passive mobile sensors. *Procedia Computer Science*, *159*, 658–667. https://doi.org/10.1016/j.procs.2019.09.221
2. Ahuja, R., & Banga, A. (2019). Mental Stress Detection in University Students using Machine Learning Algorithms. *Procedia Computer Science*, *152*, 349–353. https://doi.org/10.1016/j.procs.2019.05.007

3. Bobade, P., & Vani, M. (2020). Stress Detection with Machine Learning and Deep Learning using Multimodal Physiological Data. *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, 51–57. https://doi.org/10.1109/ICIRCA48905.2020.9183244
4. EL Bilali, A., Taleb, A., Bahlaoui, M. A., & Brouziyne, Y. (2021). An integrated approach based on Gaussian noises-based data augmentation method and AdaBoost model to predict faecal coliforms in rivers with small dataset. *Journal of Hydrology*, *599*, 126510. https://doi.org/10.1016/j.jhydrol.2021.126510
5. Ferdinando, H., Seppanen, T., & Alasaarela, E. (2016). Comparing features from ECG pattern and HRV analysis for emotion recognition system. *2016 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, 1–6. https://doi.org/10.1109/CIBCB.2016.7758108
6. Garcia-Ceja, E., Osmani, V., & Mayora, O. (2016). Automatic Stress Detection in Working Environments From Smartphones' Accelerometer Data: A First Step. *IEEE Journal of Biomedical and Health Informatics*, *20*(4), 1053–1060. https://doi.org/10.1109/JBHI.2015.2446195
7. Ghaderi, A., Frounchi, J., & Farnam, A. (2016). Machine learning-based signal processing using physiological signals for stress detection. *2015 22nd Iranian Conference on Biomedical Engineering, ICBME 2015*, 93–98. https://doi.org/10.1109/ICBME.2015.7404123
8. Gupta Megha V, & Vaikole Subhangi. (n.d.). *1*.
9. Hossain, Md. M., Asadullah, Md., Hossain, M. A., & Amin, M. S. (2022). PREDICTION OF DEPRESSION USING MACHINE LEARNING TOOLS TAKING CONSIDERATION OF OVERSAMPLING. *Malaysian Journal of Public Health Medicine*, *22*(2), 244–253. https://doi.org/10.37268/mjphm/vol.22/no.2/art.1564
10. Kumar, G. S., & Ankayarkanni, B. (2022). Comparative Study on Mental Stress Detection Using Various Stressors and Classification Techniques. *2022 International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC)*, 1–6. https://doi.org/10.1109/ASSIC55218.2022.10088365
11. Lin, H., Jia, J., Qiu, J., Zhang, Y., Shen, G., Xie, L., Tang, J., Feng, L., & Chua, T.-S. (2017). Detecting Stress Based on Social Interactions in Social Networks. *IEEE Transactions on Knowledge and Data Engineering*, *29*(9), 1820–1833. https://doi.org/10.1109/TKDE.2017.2686382
12. Mane, S. A. M., & Shinde, A. (2023). StressNet: Hybrid model of LSTM and CNN for stress detection from electroencephalogram signal (EEG). *Results in Control and Optimization*, *11*, 100231. https://doi.org/10.1016/j.rico.2023.100231
13. Mumtaz, W., Ali, S. S. A., Yasin, M. A. M., & Malik, A. S. (2018). A machine learning framework involving EEG-based functional connectivity to diagnose major depressive disorder (MDD). *Medical & Biological Engineering & Computing*, *56*(2), 233–246. https://doi.org/10.1007/s11517-017-1685-z
14. Mumuni, A., & Mumuni, F. (2022). Data augmentation: A comprehensive survey of modern approaches. *Array*, *16*, 100258. https://doi.org/10.1016/j.array.2022.100258
15. Narzary, D., & Sharma, U. (n.d.). *Brief Review on Mental Stress and Its Identification through the Machine Learning Approach*. https://ssrn.com/abstract=4031743
16. Narzary, D., Sharma, U., & Khanna, A. (2025). An automated stress detection model based on dual approach of clinical psychologist prediction and machine learning. *International Journal of Information Technology*, *17*(2), 755–765. https://doi.org/10.1007/s41870-024-02213-1
17. Nijhawan, T., Attigeri, G., & Ananthakrishna, T. (2022). Stress detection using natural language processing and machine learning over social interactions. *Journal of Big Data*, *9*(1), 33. https://doi.org/10.1186/s40537-022-00575-6
18. Onim, Md. S. H., & Thapliyal, H. (2024). Predicting Stress in Older Adults with RNN and LSTM from Time Series Sensor Data and Cortisol. *2024 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*, 300–306. https://doi.org/10.1109/ISVLSI61997.2024.00062
19. Ozsahin, D. U., Taiwo Mustapha, M., Mubarak, A. S., Said Ameen, Z., & Uzun, B. (2022). Impact of feature scaling on machine learning models for the diagnosis of diabetes. *2022 International Conference on Artificial Intelligence in Everything (AIE)*, 87–94. https://doi.org/10.1109/AIE57029.2022.00024
20. Pandey, P. S. (2017). Machine Learning and IoT for prediction and detection of stress. *2017 17th International Conference on Computational Science and Its Applications (ICCSA)*, 1–5. https://doi.org/10.1109/ICCSA.2017.8000018
21. Phutela, N., Relan, D., Gabrani, G., Kumaraguru, P., & Samuel, M. (2022). Stress Classification Using Brain Signals Based on LSTM Network. *Computational Intelligence and Neuroscience*, *2022*, 1–13. https://doi.org/10.1155/2022/7607592
22. Rosales, M. A., Bandala, A. A., Vicerra, R. R., & Dadios, E. P. (2019). Physiological-Based Smart Stress Detector using Machine Learning Algorithms. *2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)*, 1–6. https://doi.org/10.1109/HNICEM48295.2019.9073355
23. Saleh, D., Camart, N., & Romo, L. (2017). Predictors of Stress in College Students. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00019

24. Sandulescu, V., Andrews, S., Ellis, D., Bellotto, N., & Mozos, O. M. (2015). Stress detection using wearable physiological sensors. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9107*, 526–532. https://doi.org/10.1007/978-3-319-18914-7_55
25. Schmidt, P., Reiss, A., Duerichen, R., Marberger, C., & Van Laerhoven, K. (2018). Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection. *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, 400–408. https://doi.org/10.1145/3242969.3242985
26. Shanbhog M, S., & Medikonda, J. (2023). A clinical and technical methodological review on stress detection and sleep quality prediction in an academic environment. *Computer Methods and Programs in Biomedicine*, *235*, 107521. https://doi.org/10.1016/j.cmpb.2023.107521
27. Tatiur Rahman, A. K. G. Md. M. H. S. Md. M. R. (2015). Mental Stress Recognition using K-Nearest Neighbor (KNN) Classifier on EEG Signals. *ICMEIE-2015*, 1–4.
28. Winata, G. I., Kampman, O. P., & Fung, P. (2018). Attention-Based LSTM for Psychological Stress Detection from Spoken Language Using Distant Supervision. *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 6204–6208. https://doi.org/10.1109/ICASSP.2018.8461990