Effect of Anxiety Reduction Interventions on Exam Anxiety for Engineering Students Using Physiological Signals

Jyotiska Bharadwaj
Dept. of Software Engineering
Delhi Technological University
New Delhi, India
jyotiskabharadwaj@gmail.com

Ayushi Dey
Dept. of Software Engineering
Delhi Technological University
New Delhi, India
ayushidey20@gmail.com

Aditya Bibhas Sahu
Dept. of Software Engineering
Delhi Technological University
New Delhi, India
adityabibhas.sahu@gmail.com

Karan Maheshwari
Dept. of Software Engineering
Delhi Technological University
New Delhi, India
chesswithkaran@gmail.com

Sanat
Dept. of Computer Science & Engineering
Delhi Technological University
New Delhi, India
sanat23122003@gmail.com

Divyasikha Sethia
Dept. of Software Engineering
Delhi Technological University
New Delhi, India
divyashikha@dtu.ac.in

Sonia Baloni Ray
Dept. of Cognitive Science
IIIT-Delhi
New Delhi, India
sonia@iiitd.ac.in

***Abstract*—Acute test anxiety negatively affects both academic performance and mental well-being in engineering students. Although earlier studies have examined the long-term impact of culturally relevant methods like breathing techniques and Indian classical music, they mostly rely on self-reports, lack real-time physiological data, and seldom offer direct quantitative comparisons. This research investigates the short-term efficacy of guided deep breathing and Indian classical music Raga Bhairavi using wearable biosensors, statistical analysis, and machine learning. The paper randomly assigns sixty undergraduate engineering students with moderate to high anxiety to intervention and control groups. The research measures Galvanic Skin Response (GSR) and Heart Rate Variability (HRV) during the Montreal Imaging Stress Task (MIST), which induces exam-related stress. Non-parametric statistical tests confirm a significant increase in stress for the control group, while both interventions effectively reduce physiological arousal. This paper also implements various Machine learning models (Random Forest, Decision Tree, XGBoost) to classify stress states with high accuracy (>98%), F1 scores (>0.98), and ROC AUC values (>0.96). Integrating culturally grounded interventions with real-time physiological monitoring and AI-based stress classification can offer a scalable, intelligent framework to support student mental health.**

***Index Terms*—Exam Anxiety, Electrodermal Activity, Blood Volume Pulse, Indian Classical Music Therapy, Guided Breathing Intervention, Descriptive Statistics, Machine Learning in Stress Detection**

I. Introduction

Stress and anxiety are two of the most prevalent psychological issues impacting students, especially in high-pressure academic environments such as engineering institutions. Stress is typically characterized as the body's non-specific response to any demand for change, often leading to physiological and psychological alterations such as increased heart rate, elevated cortisol levels, and cognitive overload [15]. While stress can be both positive (eustress) and negative (distress), prolonged exposure to academic stress—such as intense workloads, performance pressure, and competitive examinations—often leads to anxiety, a more persistent and maladaptive emotional state [10], [13].

In general, studies divide anxiety into two categories: trait anxiety and state anxiety. State anxiety is a transient emotional state brought on by certain circumstances, such as an impending test. Trait anxiety, on the other hand, is a more consistent tendency to view circumstances as dangerous even when they are not [5]. These types of anxiety are often measured using standardized psychological tools like the State-Trait Anxiety Inventory (STAI), which captures both transient and dispositional aspects of anxiety through self-reported questionnaires [30].

Test anxiety, which combines physiological over-arousal, tension, and somatic symptoms (such as sweating, a fast heartbeat, and shortness of breath) with worry, fear of failing, and impaired cognitive functioning during or before tests, is a particularly crippling type of situational anxiety [16]. Research has found test anxiety negatively correlates with academic achievement, especially in STEM areas (science, technology, engineering, and mathematics) [10]. Additionally, it has an inverse relationship with students' psychological resilience, suggesting that kids who experience high levels of test anxiety frequently find it difficult to bounce back from academic disappointments [33].

Music therapy and guided deep breathing have gained empirical support for their physiological and psychological efficacy. Music therapy, particularly through Indian classical ragas, leverages music's emotional resonance and rhythmic structure to modulate autonomic nervous system activity. Studies have shown that listening to ragas such as Raag Hamsadhvani [20] and Raag Bhairavi [4] can lower heart rate and improve subjective well-being [8], [17].

Guided deep breathing is another effective intervention grounded in yogic and psychophysiological traditions. By regulating breath rate and depth, deep breathing activates the parasympathetic nervous system, helping to counteract the "fight-or-flight" response induced by anxiety [24], [36]. Empirical studies consistently show that deep breathing improves emotional regulation and cognitive performance. For example, Khng [24] demonstrated that primary students practising deep breathing scored better in tests and reported reduced anxiety, especially those with high baseline autonomic reactivity.

Despite their promise, few studies have directly compared these two interventions using objective physiological metrics in the context of acute test anxiety. The key research gaps are as follows:

- Most earlier research primarily relies on self-reported data, which is subject to bias and might be subjective.
- Real-time physiological monitoring is insufficient to objectively evaluate the efficacy of therapies.
- The short-term impacts of culturally relevant therapies, such as breathing exercises and Indian classical music, on engineering students are poorly studied.
- Quantitative comparisons between different intervention methods are rarely conducted.
- Limited integration of objective data analysis methods such as statistical testing and machine learning.

This study conducts a controlled experiment to compare the short-term effects of Indian raga music and guided deep breathing on undergraduate students experiencing test anxiety. After providing informed consent, 60 undergraduate students voluntarily participated in the study and were exposed to the Montreal Imaging Stress Test (MIST) [6] to induce test-related stress under controlled conditions. Wearable technology measured outcomes through psychological self-reports (STAI) and physiological signals, Heart Rate Variability(HRV), and Galvanic Skin Response (GSR). This dual-modality approach aims to identify culturally resonant, evidence-based methods for managing academic stress. Given the expanding significance of artificial intelligence and wearable health devices in healthcare, the study also looks at how they might help students' physical and mental health during stressful tests.

The study validates these short-term anxiety reduction techniques - guided deep breathing and Indian raga music - in isolation and understands how they influence both psychological perceptions and physiological responses under controlled conditions. The study outlines the following specific objectives:

- To assess the effectiveness of guided breathing and Indian classical raga music (Raga Bhairavi) in reducing test anxiety, using self-reported measures (State-Trait Anxiety Inventory) and physiological indicators (GSR, HRV).
- To examine changes in stress indicators like HRV and GSR to evaluate students' physiological reactions before and after each intervention. Following both therapies, statistically significant decreases in stress markers are confirmed by statistical analysis using the Kruskal–Wallis test, Wilcoxon signed-rank test, and Friedman's test.
- To lay the groundwork for future AI-driven systems that predict and manage anxiety levels using real-time physiological data from wearable sensors. By training machine learning models on extracted physiological features, we achieved strong performance in the binary classification of stress states. These results highlight the potential of wearable sensor-based, AI-assisted monitoring as a scalable solution for supporting student mental health.

The organization of the work is as follows. Section II covers the related work on various stress and anxiety-related studies and intervention methods. Section III presents the proposed methods for data collection, data preprocessing, statistical tests for hypothesis testing, model construction, and optimization strategies for the stress classification model. Section IV presents experimental results and discussions. Section V concludes the study and recommends future portable stress detection systems developments.

II. Related Work

Researchers have extensively studied test anxiety across various domains, with interventions such as Music Therapy, Breathing Techniques, Progressive Muscle Relaxation (PMR), and cognitive reappraisal demonstrating potential for stress reduction. This Section highlights key findings from studies evaluating anxiety-reduction techniques, particularly their effectiveness as measured by psychological and physiological parameters.

A. Music Therapy and General Anxiety

TABLE I summarizes studies that evaluated the impact of various musical interventions—from Indian classical ragas to pop music and binaural beats—on general anxiety. Notable findings indicated that Indian ragas like Bhairavi [14], Bhupali [12], and Puriya [17] significantly reduced state anxiety and improved HRV. Several studies utilized physiological markers

TABLE I
Studies on Music Therapy and General Anxiety

Ref	**Psych Test**	**Intervention**	**Sample**	**Bio-Signal**	**Highlights**
[17]	STAI	Raga Puriya, Malkauns, etc.	140 youth	sAA, Cortisol, HRV	State anxiety significantly reduced
[12]	VAS	Raga Bhupali, Pop, Silence	28 individuals	HRV	Bhupali reduced anxiety, improved vagal tone
[18]	BAI, STAI	Varying Music Tempo	21 med. students	EEG, HRV	Improved HRV, reduced anxiety
[23]	NA	Music + Yogic Breathing	400 engineering students	GSR	Flute music most effective for stress
[14]	DASS-21	Raga Bhairavi	44 patients	HRV	Reduced physiological stress

TABLE II
SUMMARY OF ANXIETY REDUCTION TECHNIQUES FOR TEST ANXIETY

Ref	Psych Test	Intervention	Sample	Bio-Signal	Highlights
[8]	BAI, WHO-5	Raga Hamsadhvani	55 medical students	None	Reduced anxiety, improved well-being
[11]	STAI, BAI, SUDS	Breathing, PMR, Dive reflex	45 OT students	None	Sensory techniques found effective
[19]	STAI	Active/Passive Music	202 pharmacy students	None	Reduced test anxiety
[9]	STAI, BP	Calm vs. Obnoxious Music	80 undergrads	BP, HR	Calm music improved scores, reduced BP and HR
[22]	STAI	Mozart Music	15 students	None	Improved trait anxiety
[7]	STAI	Music Therapy	125 nursing students	BP, vitals	Improved BP; no significant change in anxiety
[28]	Sarason's TAS	Pranayama	107 MA students	None	Reduced anxiety, enhanced performance

such as HRV, salivary alpha-amylase (sAA), and Electroencephalogram (EEG). These techniques were not explicitly designed to address test anxiety or machine learning-based prediction.

B. Studies on Test Anxiety and Interventions

This section reviews studies that targeted test anxiety among students using diverse interventions. Most studies employed psychological assessments such as the STAI, lower Beck Anxiety Inventory (BAI), and Westside Test Anxiety Scale (WTAS), while only a few incorporated physiological measures like blood pressure (BP) or heart rate (HR), as shown in TABLE II. Eyüboğlu et al. [8], Keptner et al. [11], and Husain et al. [7] test techniques including raga music (e.g., Hamsadhvani), breathing exercises, PMR, and traditional practices.

TABLE III
BREATHING TECHNIQUES AND ANXIETY

Ref	Psych Test	Intervention	Sample	Bio-Signal	Highlights
[24]	Self-reported	Deep Breathing	122 primary students	Autonomic reactivity	Reduced anxiety, improved test performance
[26]	TAS, SOS	SEL + Breathing	105 high school students	None	Improved self-efficacy and performance
[36]	FLLA, TA	Pranayama	140 undergrad learners	None	Reduced anxiety, better listening and reading
[31]	Anxiety, BP	Device-Guided Breathing	21 students	BP	Significant reduction in anxiety
[32]	Anxiety, Math	Diaphragmatic vs Gasping	103 students	None	Breathing improved math performance

Results consistently indicate reduced anxiety and enhanced well-being, though the lack of physiological data in most studies limits the objectivity of their conclusions.

C. Breathing Techniques and Anxiety

TABLE III presents studies investigating the impact of various breathing techniques—such as mindful breathing, diaphragmatic breathing, and Pranayama—on academic anxiety [36]. These interventions showed effectiveness in enhancing academic outcomes and reducing anxiety, particularly among students with heightened autonomic reactivity. However, many studies relied heavily on self-reported outcomes and small sample sizes, with few capturing objective physiological signals like HR or GSR.

III. METHODOLOGY

A. Devices and Sensors Used

This study presents a comprehensive multi-modal analysis of short-term anxiety reduction techniques in the context of exam-related stress among undergraduate engineering students. The paper monitors physiological signals using the Empatica EmbracePlus wearable device [21], [34], and a structured, controlled experimental protocol is adopted. The Empatica EmbracePlus has a Ventral Electrodermal Activity (EDA) sensor to detect subtle changes in electrical conductance at the surface of the skin and Advanced optical PPG (Photoplethysmogram) to clinically validated Pulse Rate (PR) and PR Variability (PRV) measurements through a custom-made sensor [3]. The following subsections define the signals used in the study.

1) Electrodermal Activity (EDA): EDA, also known as GSR, measures changes in the skin's electrical conductance due to sweat gland activity, which the sympathetic nervous system directly influences. This study uses EDA signals as a reliable biomarker to evaluate how well anxiety-reduction interventions work since increased skin conductance often accompanies heightened emotional states during exam-related stress.

2) Blood Volume Pulse (BVP): BVP is a physiological signal that reflects the changes in blood volume in the microvascular bed of tissue, typically measured using PPG. This study analyzes BVP signals to assess participants' physiological changes during exam-related anxiety and evaluate the impact of intervention techniques on cardiac responses.

B. Experiment Design and Data Collection

Fig. 1 outlines the six-step experimental methodology. The process begins with an extensive literature review and problem formulation. The experimental protocol is then designed and validated in consultation with domain experts. Participant recruitment follows, incorporating group-wise randomization and ensuring procedural consistency. An initial survey is conducted among the student population to collect personal information, music preferences, and self-reported anxiety levels during exams.

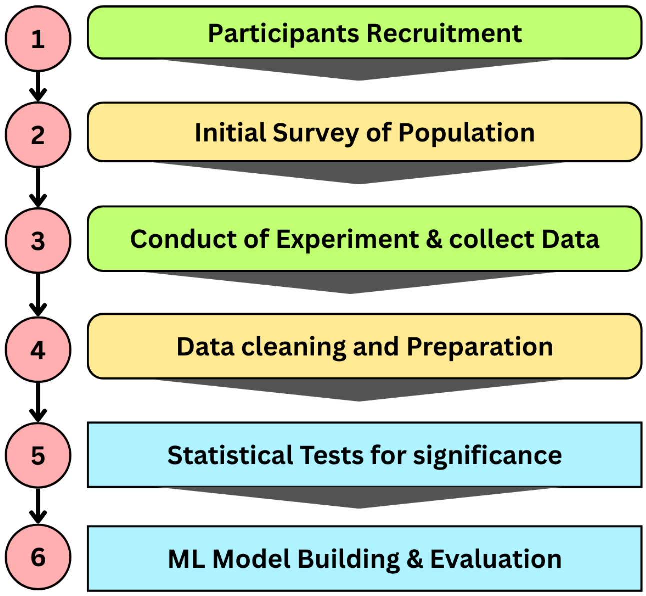

Fig. 1. Experiment design pipeline

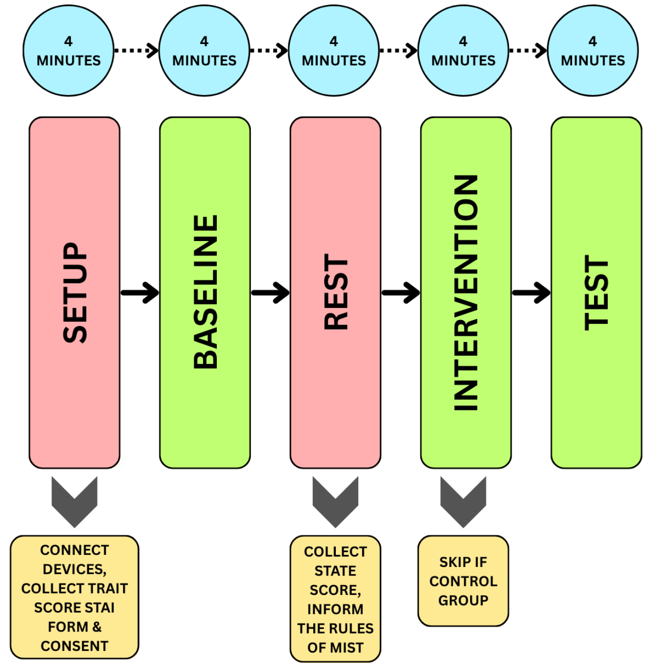

Fig. 2. Detailed experimental protocol

Participants are selected based on medium to high anxiety levels and a preference for Indian music. The STAI and the initial survey results guide the selection to ensure the relevance of the anxiety-inducing task and the effectiveness of the music-based intervention. The study obtains informed consent from all selected participants. The final participant pool consists of undergraduate engineering students in their sophomore and pre-final years at Delhi Technological University.

The main experiment follows a protocol comprising five sequential phases of four minutes each, as illustrated in Fig. 2: Setup, Baseline, Rest, Intervention, and Test. In the Setup phase, participants are connected to the Empatica EmbracePlus wearable on their non-dominant wrist. The paper reconfirms verbal consent and records trait anxiety using the STAI Trait subscale [35]. During the Baseline phase, participants sit quietly while physiological signals are recorded. In the Rest phase, participants are informed about the upcoming arithmetic test, and state anxiety scores are recorded using the STAI State subscale [35].

Depending on group allocation, participants then proceed either directly to the Test phase or through an Intervention phase:

- *Group A (Control):* No anxiety-reduction intervention; participants proceed directly to the test phase.
- *Group B (Guided Breathing):* Participants undergo a 4-minute guided breathing session using a medically approved audio track [29].
- *Group C (Raga Bhairavi):* Participants listen to a 4-minute instrumental flute rendition of Raga Bhairavi, chosen for its calming properties and the absence of lyrics to minimize cultural bias [1].

In the Test phase, all participants perform the Montreal Imaging Stress Test (MIST), a time-constrained arithmetic task known to induce cognitive stress [27]. The MIST is widely used in neuropsychological studies and is well-suited for experiments involving physiological signal monitoring [2].

Continuous physiological data, including EDA, BVP, and HR, is collected throughout the experiment. The STAI assesses trait and state anxiety, with the trait score recorded during setup and the state score just before the test.

Following data collection, this study preprocesses the signals and performs statistical analyses. The study then develops and evaluates machine learning models to analyze anxiety responses and the effectiveness of the intervention.

C. Data Preprocessing and Preparation

This work applies a rigorous preprocessing pipeline to the physiological signals to enhance signal quality and analytic reliability.

The study uses z-score normalization for EDA signals to standardize amplitudes and reduce inter-subject variability. Each time series $X = \{x_1, x_2, \ldots, x_n\}$ is transformed using:

$$z_i = \frac{x_i - \mu}{\sigma} \tag{1}$$

Where μ and σ represent the mean and standard deviation for the given participant. This normalization ensures all participants' signals are on the same scale (see Equation 1).

The study applies median baseline correction for BVP signals to eliminate low-frequency drifts and outliers. The raw signal $S = \{s_1, s_2, \ldots, s_n\}$ is passed through a median filter with window size w, yielding a baseline estimate $B = \text{median\_filter}(S, w)$. Equation 2 computes the corrected signal and improves downstream HRV feature extraction by stabilizing the baseline.

$$s_i^{(\text{corrected})} = s_i - B_i \tag{2}$$

The study performs artefact detection on BVP data using the Kubios HRV algorithm (via the NeuroKit2 library), which flags implausible inter-beat intervals (IBIs). For a sequence $\{\text{IBI}_1, \text{IBI}_2, \ldots, \text{IBI}_n\}$, segments are flagged if:

$$|\text{IBI}_i - \text{IBI}_{i-1}| > \theta \tag{3}$$

where θ is a physiologically-informed threshold. Depending on severity and density, the researchers interpolate or remove flagged segments, as defined in Equation 3.

The study segments signals by phase—Setup, Baseline, Rest, Intervention (if applicable), and Test (MIST)—each lasting four minutes. Phase labels are assigned using synchronized timestamps. This division ensures that all signals align with the study protocol and are artefact-free, enabling robust statistical analysis and model development.

D. Feature Extraction

After preprocessing and phase-wise segmentation of the physiological signals, the study extracts time-domain features from EDA and BVP signals for each experimental phase. The selected features capture both tonic and phasic characteristics of sympathetic activity (from EDA) and HRV and cardiovascular dynamics (from BVP-derived inter-beat intervals). These features are widely used in psychophysiological studies to assess autonomic nervous system responses to stress and relaxation interventions. For EDA, features include the mean and standard deviation of Skin Conductance Levels (SCL), the number of Skin Conductance Responses (SCR counts), and the mean amplitude of the phasic component. For BVP-derived HRV features, the study computes the time-domain metrics such as SDNN, RMSSD, mean HR, pNN50, and Poincaré plot descriptors (SD1 and SD2) using the NeuroKit2 library [25] and validates the physiological algorithms. TABLE IV summarises all extracted features and their physiological interpretation. These features serve as quantitative indicators of stress and relaxation responses and are used in subsequent statistical analyses and machine learning modelling to evaluate intervention effectiveness and predict anxiety states.

E. Selection of Data Analysis Methods

Based on the features of the data, the study chooses a set of tests to evaluate the statistical significance of physiological changes across various experimental circumstances and phases. The research starts with exploratory data visualization using error plots of the collected characteristics to evaluate patterns and variability between groups. Fig. 3 shows a rise in HR signal in Group A (Control), but it does not indicate such changes in the case of Groups B and C (Intervention).

TABLE IV
EXTRACTED PHYSIOLOGICAL FEATURES

Feature Name	Signal Source	Description
scl_mean	EDA	Mean Skin Conductance Level (tonic component)
scl_std	EDA	Standard deviation of SCL indicating variability in tonic arousal
scr_counts	EDA	Number of discrete Skin Conductance Responses (phasic arousal events)
phasic_mean	EDA	Mean amplitude of phasic responses
sdnn	BVP/HRV	Standard Deviation of NN intervals (overall HRV)
rmssd	BVP/HRV	Root Mean Square of Successive Differences (parasympathetic activity)
hr	BVP	Average heart rate during the phase
pnn50	BVP/HRV	Percentage of successive NN intervals differing by more than 50 ms
poincare_sd1	BVP/HRV	Short-term HRV from Poincaré plot (parasympathetic modulation)
poincare_sd2	BVP/HRV	Long-term HRV from Poincaré plot (sympathetic-parasympathetic balance)

TABLE V
NORMALITY TEST (SHAPIRO-WILK)

Metric	Group A (Control)	Group B (Raga)	Group C (Breathing)
scl_mean	W=0.786, p=0.001	W=0.908, p=0.059	W=0.921, p=0.116
scl_std	W=0.831, p=0.003	W=0.980, p=0.938	W=0.911, p=0.077
phasic_mean	W=0.841, p=0.004	W=0.947, p=0.328	W=0.948, p=0.369
scr_count	W=0.938, p=0.220	W=0.788, p=0.001	W=0.940, p=0.267
SDNN	W=0.419, p=0.000	W=0.902, p=0.046	W=0.259, p=0.000
RMSSD	W=0.448, p=0.000	W=0.863, p=0.009	W=0.264, p=0.000
HR	W=0.932, p=0.168	W=0.714, p=0.000	W=0.687, p=0.000
pNN50	W=0.901, p=0.044	W=0.890, p=0.026	W=0.965, p=0.682
Poincare_SD1	W=0.448, p=0.000	W=0.863, p=0.009	W=0.263, p=0.000
Poincare_SD2	W=0.394, p=0.000	W=0.932, p=0.170	W=0.257, p=0.000

1) Shapiro-Wilk Test: This test determines if a sample comes from a regularly distributed population. It provides a p-value to indicate the degree of normality and works particularly well with small sample numbers. Following extracting relevant physiological features, as outlined in TABLE IV, the study evaluates the dataset for normality using the Shapiro-Wilk test. Results reveal that the majority of features across the three groups (Raga, Breathing, Control) do not follow a normal distribution (TABLE V) (not normally distributed if p-value <0.05). The paper uses the following non-parametric tests to ensure robustness against violations of normality assumptions.

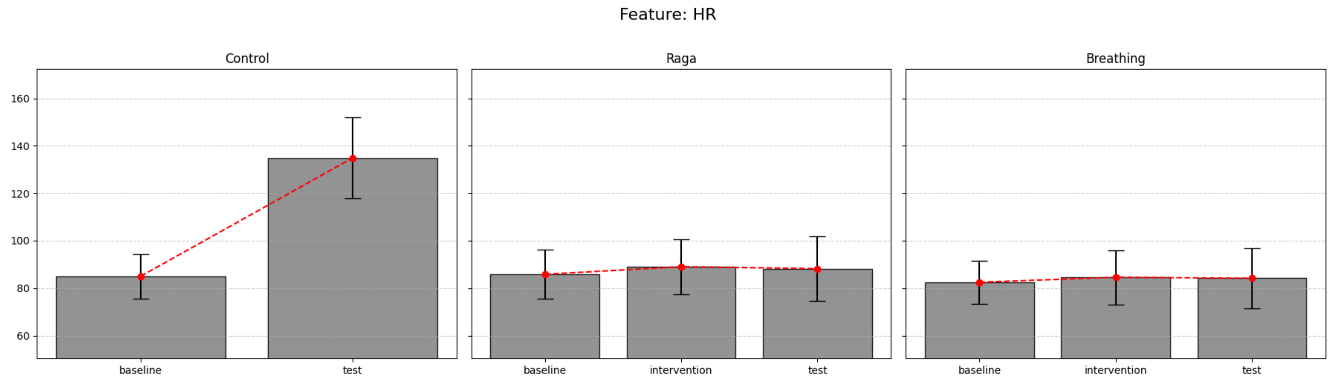

Fig. 3. Error plots of Heart Rate Signal

2) Friedman's test: This non-parametric statistical test can identify variations in treatment throughout several test attempts. When there are no assumptions of parametric tests, it is a substitute for repeated measures of the Analysis Of Variance (ANOVA) technique.

3) Wilcoxon Signed-Rank Test: This non-parametric statistical test compares two related samples or repeated measurements on a single sample. It is a non-parametric substitute for the paired t-test. It determines whether their population mean ranks vary—the test works by ranking the absolute differences between paired observations and examining the signed-rank distribution. When the differences between pairs are often not dispersed, it is advantageous.

4) Kruskal-Wallis Test: This non-parametric substitute for one-way ANOVA determines whether there are statistically significant differences between the medians of three or more independent groups. It is particularly suitable in cases with no normality and homogeneity of variances assumptions of ANOVA. After ranking every observation across groups, the approach determines whether there are any notable differences in the average ranks.

Specifically, the Kruskal-Wallis H test and Wilcoxon Signed Rank Test compare independent groups, and the Friedman test is applied to detect within-subject changes across repeated measures. The following analyses address the research objectives systematically:

- *Analysis 1 (Baseline comparison for all groups):* The study applies Friedman's test to verify that the initial (baseline) conditions across all groups are statistically comparable.
- *Analysis 2 (Stress Induction in Control Group):* The study next evaluates whether physiological features in the test phase of the control group differ significantly from the baseline, indicating whether the MIST test successfully induces stress in the absence of intervention. The study applies the Wilcoxon signed-rank test (within-group) to compare physiological metrics from the baseline and test phases.
- *Analysis 3 (Stress Induction in Intervention Groups):* For the Raga and Guided Breathing groups, the research again applies the Friedman test to compare baseline, intervention, and test phase features within each group. This experiment compares pre- and post-intervention physiological responses to whether the test phase induces stress despite the intervention.
- *Analysis 4 (Effect of Intervention):* To evaluate the effectiveness of the interventions, the study performs the Kruskal-Wallis test (between-group) to compare the test phase features of the intervention groups (Raga and Breathing) against the control group. The study calculates a delta value as (feature value in test phase - feature value in baseline). A statistically significant difference here would suggest that the interventions modulate participants' stress responses during the MIST phase.

F. Machine Learning Model Building and Optimizations

To evaluate the potential of physiological features in classifying anxiety states, the study develops several supervised machine learning models to distinguish between high-stress (test phase within the control group) and low-stress (baseline, test, and intervention phases within Raga and Breathing groups) conditions. The models include Support Vector Machines with radial basis function (RBF) and polynomial kernels, Decision Tree, Random Forest, XGBoost, K-nearest neighbours (KNN), Naive Bayes, and Logistic Regression. Fig. 4 describes the model-building process. The following Subsections from III-F1 to III-F6 describe the steps involved.

1) Data Preparation and Preprocessing: To ensure data integrity, the researchers exclude incomplete or corrupted records from analysis. Non-predictive variables, such as participant identifiers, timestamps, and protocol metadata, are removed to prevent redundancy and potential data leakage. The study retains only numerical features to ensure compatibility with ML algorithms and downstream resampling techniques. To mitigate ordering bias, the study randomly shuffles the dataset. Z-score normalization via the Python library StandardScaler standardizes feature values implemented within a pipeline to ensure consistent preprocessing across all models.

2) Label Assignment: The study assigns binary labels to reflect two distinct stress states based on the results of the statistical tests. Section IV-B covers the details of these labels.

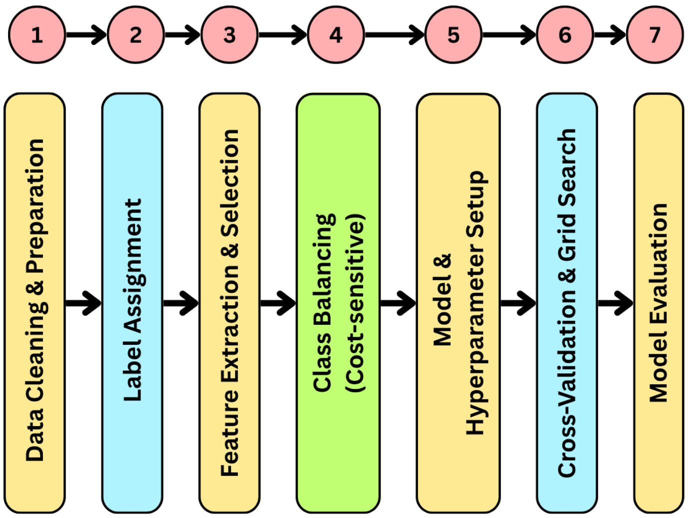

Fig. 4. Machine learning model building pipeline

3) Feature Selection: The study excludes all non-informative and potentially confounding attributes, such as group labels and phase indicators. The final feature set comprises only numerical, physiologically relevant variables (as defined in TABLE IV), facilitating compatibility with techniques like SMOTE and reducing the risk of model bias from high-level descriptive features.

4) Handling Class Imbalance: Given the inherent class imbalance—where low-stress samples significantly outnumber high-stress ones—the researchers evaluate multiple strategies, including random undersampling and oversampling. However, these approaches lead to either information loss or overfitting. Ultimately, the study employs cost-sensitive learning by adjusting the classification loss function to impose greater penalties on minority class misclassifications. This algorithm-level solution preserves the original data distribution while improving model sensitivity and robustness.

5) Model Selection and Optimization: The study evaluates a suite of supervised classification algorithms, including Logistic Regression, Support Vector Machines (RBF and polynomial kernels), Decision Trees, Random Forests, K-Nearest Neighbors, Naive Bayes, and XGBoost. A pipeline that includes preprocessing steps encapsulates each model. Grid search combined with stratified 5-fold cross-validation tunes the hyperparameter tuning to maintain class balance across folds and ensure reliable model evaluation. Optimization targets both accuracy and AUC-ROC as primary performance criteria.

6) Performance Evaluation: The study assesses model performance using a set of standard classification metrics, each providing unique insights into how well the model identifies high-stress instances.

- *True Positive (TP):* A high-stress case is accurately predicted by the model to be high-stress.
- *True Negative (TN):* The model accurately predicts low stress as low stress.
- *False Positive (FP):* The model mistakenly predicts A low-stress event as high-stress.
- *False Negative (FN):* The model mistakenly predicts A high-stress event as low-stress.

Based on these values, the study defines the following evaluation metrics:

- *Accuracy:* Measures the overall correctness of the model's predictions.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

- *Precision:* Indicates how many instances predicted as high-stress are high-stress.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{5}$$

- *Recall (Sensitivity):* Measures the model's ability to correctly identify high-stress cases.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{6}$$

- *F1-Score:* Harmonic mean of Precision and Recall, useful for imbalanced class distributions.

$$\text{F1-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{7}$$

- *AUC-ROC:* Area Under the Receiver Operating Characteristic Curve; this metric summarizes the trade-off between the actual positive rate (Recall) and the false positive rate across various thresholds. A higher AUC indicates better model discrimination capability.

7) Model Interpretation and Comparative Analysis: The study conducts a comparative analysis across classifiers and class-balancing strategies. The paper evaluates the models trained using cost-sensitive learning alongside those trained on resampled datasets (via undersampling, oversampling, or SMOTE). Performance differences are analyzed to assess trade-offs between predictive accuracy, sensitivity to minority classes, and generalizability. This analysis provides insights into the effectiveness of algorithm-level versus data-level imbalance handling techniques in psychophysiological stress classification tasks.

IV. Results and Discussion

A. Statistical Results

1) Analysis 1 (Baseline comparison for all groups): The results, presented in TABLE VI, reveal no significant differences in baseline physiological metrics across the three groups (p-value >0.05 for most of the features). This result confirms that the random group allocation does not introduce any systematic bias at the start of the experiment.

2) Analysis 2 (Stress Induction in Control Group): TABLE VII) compares physiological metrics from the baseline and test phases. Results show a statistically significant increase in stress markers within the control group, indicating successful stress induction by MIST. P-value <0.05 signifies a significant difference between each group's baseline and test phases.

3) Analysis 3 (Stress Induction in Intervention Groups): TABLE VIII indicates no significant differences in most metrics, suggesting that physiological stress levels remain statistically stable throughout the experimental phases in both the Raga and Breathing groups. (p-value >=0.05 signifies that there is not any significant difference between the baseline and test phases of each group) These findings support the conclusion that short-duration interventions effectively buffer against acute stress responses, unlike the control group.

4) Analysis 4 (Effect of Intervention): TABLE IX compares delta values of the physiological metrics from the baseline and test phases. A p-value <0.05 signifies that the delta values (value in test phase - value in baseline) of the control group and the intervention group have significant differences, suggesting a rise in stress markers in the test phase of the control group but not in intervention groups. Results show a statistically significant increase in stress markers within the control group, indicating successful stress induction. In contrast, no such elevation is observed in the Raga or Breathing groups, suggesting that both interventions had a stabilizing or protective effect.

These results align with the study's core objective: to evaluate the short-term efficacy of culturally resonant (Indian raga) and evidence-based (guided breathing) techniques in mitigating acute test anxiety. The absence of significant stress elevation post-intervention in both experimental groups supports the potential of these techniques for real-time stress regulation. Although the results are encouraging, it is crucial to recognize that this study only looks at short-term effects. Future studies should investigate the long-term effects of repeated exposure to these therapies. Such strategies could considerably lessen exam anxiety and buffer against situational stressors with regular practice.

B. Machine Learning Analysis

This Section evaluates and compares the performance of eight machine learning models in classifying stress states using physiological features. The study assigns binary labels to reflect two distinct stress states, informed by results from Section IV-A:

- *Label 0 (Low Stress):* From the results in SubsectionIV, it is concluded that all phases (Baseline, Intervention, and Test) within the Raga and Guided Breathing groups are statistically similar.
- *Label 1 (High Stress):* Test phase of the Control group.

TABLE VII
WILCOXON TEST P-VALUES ACROSS GROUPS

Metric	Raga p-value	Breathing p-value	Control p-value
scl_mean	0.812355	0.028931	0.004860
scl_std	0.474905	0.515278	0.000002
phasic_mean	0.008308	0.984322	0.311794
scr_count	0.394854	0.525658	0.000002
signal_quality	0.781511	0.317311	0.098134
SDNN	0.001927	0.952984	0.000002
RMSSD	0.043620	1.000000	0.000002
HR	0.525292	0.629128	0.000002
pNN50	0.001151	0.434756	0.000002
Poincare_SD1	0.043620	1.000000	0.000002
Poincare_SD2	0.000771	0.594887	0.000002

Statistical analyses support these assignments, which show (i) no significant physiological differences across baselines of all groups, (ii) significant stress elevation in the Control group's test phase, and (iii) no significant change in physiological states across phases in the intervention groups. As a result, the study labels only the Control test phase as a high-stress condition, validating a binary classification framework.

All models exhibit strong classification capabilities, with several achieving high accuracy and robustness across multiple evaluation metrics. These results confirm that features derived from GSR and HRV signals effectively distinguish between high and low-stress states when combined with appropriate preprocessing and class-balancing strategies. Consistently high scores across metrics such as accuracy, F1 score, precision, recall, and ROC AUC affirm the extracted features' reliability and the modelling pipeline's strength. These findings underscore the potential of automated, physiological-signal-based stress detection in academic or high-pressure environments.

Moreover, the machine learning results reinforce the behavioural findings. The models successfully classify the test-phase data from the control group as "stressed" while correctly identifying test-phase samples from the Raga and Breathing groups as "non-stressed." This result suggests that both interventions produce measurable physiological changes that

TABLE VI
FRIEDMAN TEST: BASELINE COMPARISON ACROSS GROUPS

Metric	Statistic	p-value
scl_mean	0.315789	0.853940
scl_std	0.315789	0.853940
phasic_mean	8.000000	0.018316
scr_count	1.972973	0.372885
signal_quality	0.291667	0.864302
SDNN	2.947368	0.229080
RMSSD	2.000000	0.367879
HR	2.210526	0.331124
pNN50	0.421053	0.810158
Poincare_SD1	2.000000	0.367879
Poincare_SD2	3.263158	0.195620

TABLE VIII
FRIEDMAN TEST RESULTS: RAGA AND BREATHING

Metric	Raga Statistic	Raga p-value	Breathing Statistic	Breathing p-value
scl_mean	4.900	0.086294	6.000	0.049787
scl_std	3.700	0.157237	0.105	0.948729
phasic_mean	4.300	0.116484	5.474	0.064775
scr_count	4.081	0.129958	1.380	0.501505
SDNN	7.300	0.025991	2.000	0.367879
RMSSD	6.100	0.047359	2.000	0.367879
HR	1.900	0.386741	4.526	0.104021
pNN50	7.600	0.022371	0.108	0.947381
Poincare_SD1	6.100	0.047359	2.000	0.367879
Poincare_SD2	7.300	0.025991	4.526	0.104021

TABLE IX
KRUSKAL-WALLIS TEST ON DELTA FEATURES (TEST - BASELINE)

Metric	Raga H	Raga p-value	Breathing H	Breathing p-value
scl_mean	1.42e-14	0.999999	12.534	0.0003997
scl_std	22.924	0.000001686	25.864	0.0000003663
phasic_mean	4.567	0.03260149	0.967	0.3254032
scr_count	15.837	0.00006905	23.396	0.000001318
SDNN	28.713	0.00000008395	25.864	0.0000003663
RMSSD	26.162	0.0000003139	22.816	0.000001783
HR	29.004	0.00000007225	28.503	0.00000009356
pNN50	28.423	0.00000009748	22.284	0.000002351
Poincare_SD1	26.162	0.0000003139	22.816	0.000001783
Poincare_SD2	29.004	0.00000007225	27.313	0.0000001731

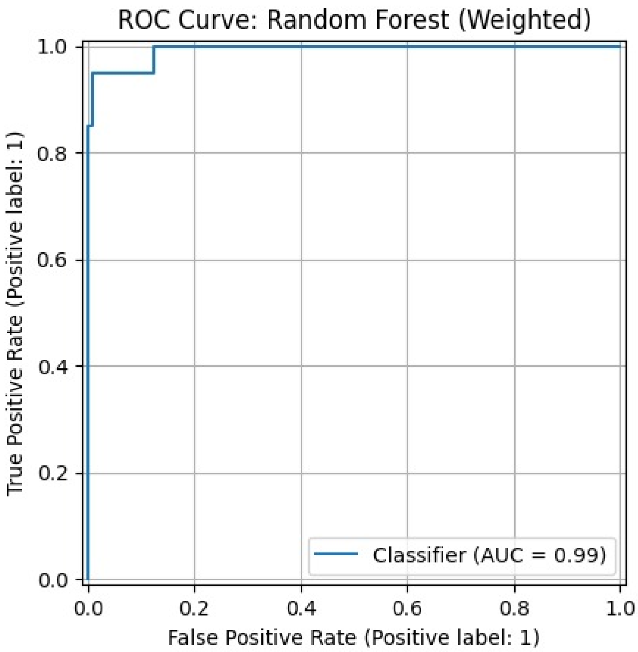

Fig. 5. ROC-AUC of Random Forest

reduce stress expression, thereby enhancing the practical relevance of these techniques. Integrating cultural and evidence-based interventions with machine learning provides a promising framework for accessible and scalable stress regulation strategies.

TABLE X summarizes the performance of all models across different evaluation metrics. Decision Tree, Random

TABLE X
COMPARATIVE PERFORMANCE METRICS (COST-SENSITIVE LEARNING)

Model	Accuracy	F1-Score	ROC AUC
Kernel SVM (RBF)	0.968354	0.967242	0.994203
Kernel SVM (Poly)	0.962025	0.960187	0.883333
Decision Tree	0.981013	0.981210	0.962862
Random Forest	0.981013	0.980804	0.993116
XGBoost	0.981013	0.980345	0.984964
KNN	0.962025	0.959079	0.898913
Naive Bayes	0.677215	0.731234	0.942391
Logistic Regression	0.911392	0.918588	0.924275

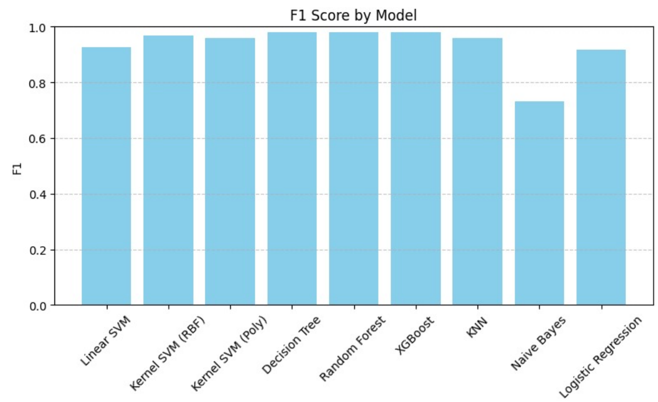

Fig. 6. F1 score comparison

Forest, and XGBoost exhibit the strongest and most consistent performance among the classifiers. All three achieve an accuracy of 0.981, with F1 scores above 0.98 and ROC AUC values exceeding 0.96. Fig. 6 compares the F1 scores across models. In particular, XGBoost balances precision and recall exceptionally well, offering high discrimination power without overfitting. Fig. 5 presents the Random Forest model's ROC curve and AUC values. The study demonstrates that machine learning models—primarily ensemble and kernel-based methods—can effectively classify exam-induced stress using physiological signals. These results underscore the feasibility of integrating sensor-based monitoring with culturally contextual interventions like Indian classical music to enhance student well-being.

V. CONCLUSION AND FUTURE WORK

This study investigates the short-term effectiveness of two non-pharmacological interventions—guided deep breathing and Indian classical Raga Bhairavi music—in alleviating exam-related anxiety among engineering students. Statistical analysis of physiological markers (EDA and HRV) reveals that while the control group exhibits significant stress responses, both intervention groups show no such increase. The study trains machine learning models on wearable sensor data. It demonstrates their ability to effectively distinguish between high- and low-stress states, reinforcing the feasibility of real-time stress monitoring. These findings support the integration of culturally resonant, low-cost interventions into academic wellness programs and highlight the potential of sensor-based, ML-driven mental health support systems. The work bridges psychology, physiology, and data science by combining objective measurements with machine learning to assess emotional regulation. However, the study has limitations. The study only looks at short-term analysis; the sample is small and uniform (N=60). It conducts the experiments in a controlled laboratory setting, which might not accurately represent academic stress in the real world. Furthermore, it is still difficult to comprehend complicated machine learning models, and physiological information obtained from wearables is frequently noisy.

Building on these promising results, future studies can examine the long-term efficacy of these interventions by evaluating cumulative effects on students' baseline anxiety and academic performance through longitudinal research. The research also sees potential in developing personalized stress management systems where machine learning models adapt in real-time to individual physiological responses and recommend tailored interventions such as specific ragas or breathing techniques. Furthermore, deploying these interventions in real-world educational environments—such as classrooms or examination halls—would enhance ecological validity and assess integration feasibility within institutional mental health programs. Including additional biosignals like EEG or respiration rate could support a more comprehensive modelling of stress and emotional states, thereby improving prediction accuracy. Finally, mobile or wearable-based applications that deliver automated interventions and provide real-time biofeedback could democratize access to mental wellness support, especially for students in high-pressure academic settings.

References

[1] Hariprasad Chaurasia. Raag bhairavi. https://www.youtube.com/watch?v=tLXNNejKhJs, 2023. Accessed: 2025-06-20.

[2] Katarina et al. Dedovic. The montreal imaging stress task: Using functional imaging to investigate the effects of perceiving and processing psychosocial stress in the human brain. *Journal of Psychiatry and Neuroscience*, 30(5):319–325, 2005.

[3] Empatica Inc. EmbracePlus: the world's most advanced smartwatch for continuous health monitoring. https://www.empatica.com/en-int/embraceplus/, 2025. Accessed: 2025-06-20.

[4] Chandra et al. Impact of indian classical raga in immersive environments on human psycho-physiological parameters. 06 2023.

[5] Charles D. Spielberger et al. State-trait anxiety inventory, 1964. Accessed: 2024-12-11.

[6] De Calheiros Velozo et al. The repeated Montreal Imaging Stress Test (rmist): Testing habituation, sensitization, and anticipation effects to repeated stress induction. *Psychoneuroendocrinology*, 128:105217, June 2021. Epub 2021 Apr 18. PMID: 33882371.

[7] Gülcan Eyüboğlu et al. Effect Of Music Therapy On Nursing Students' First Objective Structured Clinical Exams, Anxiety Levels And Vital Signs: A Randomized Controlled Study. *Nurse Education Today*, 97, 2021.

[8] Husain et al. Effect of Receptive Music Therapy on Exam-related Anxiety in Postgraduate Medical Students. *Int J of Medical Science*, 9(11):243–248, 2021.

[9] Jennifer L. Lilley et al. Effects Of Music And Grade Consequences On Test Anxiety And Performance. *Psychomusicology: Music, Mind, and Brain*, 24:184–190, 2014.

[10] Justin Charles Major et al. Effects Of Test Anxiety On Engineering Students' STEM Success. In *2020 ASEE Virtual Annual Conference Content Access*, 2020.

[11] Karen M. Keptner et al. Effectiveness Of Anxiety Reduction Interventions On Test Anxiety: A Comparison Of Four Techniques Incorporating Sensory Modulation. *British Journal of Occupational Therapy*, 84:289–297, 2021.

[12] Karuna Nagarajan et al. Immediate Effect Of Indian Music On Cardiac Autonomic Control And Anxiety: A Comparative Study. *Heart India*, 3:93–100, 2015.

[13] Keptner et al. Effectiveness of Anxiety Reduction Interventions on Test Anxiety: A Comparison of Four Techniques Incorporating Sensory Modulation. *British Journal of Occupational Therapy*, 9(11):243–248, 2020.

[14] Kulbhushan Chand et al. Raga Bhairavi In Virtual Reality Reduces Stress-related Psychophysiological Markers. *Scientific Reports*, 14, 2024.

[15] Laura P. Jiménez-Mijangos et al. Advances And Challenges In The Detection Of Academic Stress And Anxiety In The Classroom: A Literature Review And Recommendations. *Education and Information Technologies*, 28:3637–3666, 2023.

[16] Pekrun et al. Test Anxiety and Academic Achievement. *International Encyclopedia of the Social Behavioral Sciences*, 9(11):244–249, 2015.

[17] Santosh K. Sharma et al. Effect Of Indian Music As An Auditory Stimulus On Physiological Measures Of Stress, Anxiety, Cardiovascular And Autonomic Responses In Humans—A Randomized Study. *Journal of Evidence Based Integrative Medicine*, 26:365–375, 2021.

[18] Sushma Sharma et al. Indian Classical Music With Incremental Variation In Tempo And Octave Promotes Better Anxiety Reduction And Controlled Mind Wandering—A Randomised Controlled EEG Study. *EXPLORE: The Journal of Science and Healing*, 17:115–121, 2020.

[19] Suzanne Galal et al. Effectiveness Of Music Interventions To Reduce Test Anxiety In Pharmacy Students. *Pharmacy*, 9:10, 2021.

[20] Vaishali C V et al. Effects of Hamsadhwani music on heart rate variability: Gender discrepancies. *International Journal of Biomedical Research*, 11(3):e5386, 2020.

[21] Giulia Gerboni, Giulia Comunale, Weixuan Chen, Jessie Lever Taylor, Matteo Migliorini, Rosalind Picard, Marisa Cruz, and Giulia Regalia. Prospective clinical validation of the Empatica Embraceplus wristband as a reflective pulse oximeter. *Frontiers in Digital Health*, 5:1258915, 2023.

[22] PeiWen Huang. Impact Of Classical Piano Music On Student Test Anxiety. *Advances in Social Science, Education and Humanities Research*, 216:1887–1895, 2023.

[23] Anurag Joshi and Ravi Kiran. Gauging The Effectiveness Of Music And Yoga For Reducing Stress Among Engineering Students: An Investigation Based On Galvanic Skin Response. *Work*, 65(3):671–678, 2020.

[24] Kiat Hui Khng. A better state-of-mind: Deep breathing reduces state anxiety and enhances test performance through regulating test cognitions in children. *Cognition and Emotion*, 31(7):1502–1510, 2017.

[25] Dominique Makowski, Tam Pham, Zen J. Lau, Jan C. Brammer, François Lespinasse, Hung Pham, Christopher Schölzel, and S. H. Annabel Chen. NeuroKit2: A python toolbox for neurophysiological signal processing. *Behavior Research Methods*, 53(4):1689–1696, feb 2021.

[26] Carolyn et al. McLeod. The effectiveness of social-emotional learning strategies and mindful breathing with biofeedback on the reduction of adolescent test anxiety. *Canadian Journal of Education*, 44(3):815–847, 2021.

[27] Millisecond Software, LLC. Montreal stress test. https://www.millisecond.com/download/library/montrealstresstest, 2024. Accessed: 2024-12-13.

[28] Azadeh Nemati. The Effect Of Pranayama On Test Anxiety And Test Performance. *International journal of yoga*, 6(1):55–60, 2013.

[29] Indiana University School of Medicine. Belly breathing exercise: Calming the body and mind. https://www.youtube.com/watch?v=7Ep5mKuRmAA, 2023. Accessed on: June 20, 2025.

[30] OptimistMinds. State-trait anxiety inventory. https://optimistminds.com/state-trait-anxiety-inventory/, 2023. Accessed: 2024-12-11.

[31] Zehava et al. Ovadia-Blechman. Reducing test anxiety by device-guided breathing: A pilot study. *Frontiers in Psychology*, 13:678098, 2022.

[32] Erik et al. Peper. Breathing and math performance: Implications for performance and neurotherapy. *NeuroRegulation*, 3(4):142–142, 2016.

[33] Anil Jose PS and Devika BS. Test Anxiety And Resilience Among Engineering Students. *Open Access Journal of Behavioural Science & Psychology*, 7, 2024.

[34] Giulia Regalia, Francesco Onorati, Matteo Lai, Chiara Caborni, and Rosalind W Picard. Multimodal wrist-worn devices for seizure detection and advancing research: focus on the Empatica wristbands. *Epilepsy research*, 153:79–82, 2019.

[35] Charles D. Spielberger. *State-Trait Anxiety Inventory for Adults: Manual and Sample: Manual, Instrument and Scoring Guide*. Consulting Psychologists Press, Palo Alto, CA, 1983.

[36] Murselin et al. Tasan. The effect of pranayamic breathing as a positive psychology exercise on foreign language learning anxiety and test anxiety among language learners at tertiary level. *Frontiers in Psychology*, 12:742060, 2021.