Machine Learning based Anxiety Detection using Physiological Signals and Context Features

Arushi Jain
Dept. of Computer Science and Engineering
National Institute of Technology, Hamirpur
Hamirpur, India
arushijain0202@gmail.com

Rajeev Kumar
Dept. of Computer Science and Engineering
National Institute of Technology, Hamirpur
Hamirpur, India
rajeev@nith.ac.in

Abstract—Anxiety is a common mental disorder that affects millions of people worldwide. Anxiety can be detected using physiological signals such as heart rate, skin conductance, blood pressure, and respiration. Machine learning is the ability to learn from data and forecast future events. Physiological signals can be analyzed and classified into different anxiety levels using machine learning. This paper aims to present the contribution of context features in detection of anxiety using physiological signals via machine learning. It is an imbalanced multi-classification problem using ECG and EDA signals with context features collected from 15 healthy individuals. These signals of sampling rate 700 Hz collected using RespiBAN from the Wearable Stress and Affect Detection (WESAD) dataset are used here with the 6-STAI questionnaire and the context collected about the subject during the experiment. Scores for each subject are calculated from this questionnaire which categorized subjects into different levels of anxiety. Multiple features are extracted from these signals and scaling is done on the extracted features. With the help of different data balancing techniques and different machine learning algorithms, this study will classify anxiety into three classes, namely, low, moderate and high. Classification of anxiety with fusion of multimodal physiological signals gave better results as compared to single modal physiological signals with an accuracy of 89.8% with ECG signal, 85.9% with EDA signal and 96.7% with ECG and EDA both using Gradient Tree Boosting algorithm with LOOCV after random oversampling. While, the best results are achieved when multimodal physiological signals are fused with context features, it gives the accuracy of 97.3% with Gradient Tree Boosting and LOOCV algorithm.

Keywords—anxiety detection, physiological signals, ECG, EDA, rr-interval, data balancing techniques, context features.

I. INTRODUCTION

Stress and anxiety can have negative impacts on people's physical and mental well-being, as well as their social and occupational functioning. Anxiety disorders are the most prevalent mental illnesses worldwide, impacting 301 million individuals in 2019, according to the WHO. There has been a marked rise in the prevalence of affective disorders during the COVID-19 pandemic, including a 27% increase in depressive disorders and a 25% increase in anxiety disorders [1]. Prolonged anxiety can have harmful effects on mental as well as physical health which includes increased risk of heart diseases, heart attack, high blood pressure, sudden weight gain or loss, sleep problems, etc. It can impair cognitive abilities, such as attention, concentration, learning and recall.

Anxiety is a complex phenomenon that involves both psychological and physical aspects. Physiological signals are the measurable changes in the body that occur when a person experiences anxiety. Some of the physiological signals of interest for anxiety detection are hormone levels, electrocardiogram (ECG), electrodermal activity (EDA), electroencephalogram (EEG), blood pressure (BP), galvanic skin response (GSR), electromyogram (EMG) , respiration volume, and pupil diameter [2,3].

II. RELATED WORK

In order to determine anxiety and stress on the basis of their physiological features, a number of machine learning algorithms have been applied. Many studies have been done on stress and anxiety classification using single signals as well as multiple signals with different classes of classification including, binary (anxious or non-anxious), three classes (low, moderate and high) or four classes (low, mild, moderate and high).

Anxiety classification has been done using an ECG signal [6]. For the extraction of features, the proposed system uses the Pan Tompkins algorithm, Support Vector Machine (SVM) and Kalman filtering for classification, and QRS detection. The classification of SVMs uses specific parameters, e.g. the mean RR intervals, and minimum and maximum RR interval values. To identify arousal associated with anxiety, an unsupervised algorithm called Kalman filtering is used which uses the state dynamics of RR intervals. Kalman filters incorporate additive white Gaussian noise into a linear model of the RR interval. In all respects, the system is based on sophisticated algorithms and physiological measures to accurately identify anxiety. The system can identify anxiety states with 71.1% and 69.5% accuracy using the SVM classifier and Kalman filter, respectively.

Classification of anxiety using multiple signals like ECG, EDA and EMG has also been proposed [7]. In this study, classification of anxiety has been seen as an imbalanced binary classification problem. It employed unsupervised and supervised feature selection techniques which revealed that ECG features play an important role in anxiety classification.

A study emphasized the possible benefits of integrating mental stress detection with Virtual Reality Exposure Therapy (VRET) [8]. The four levels of the anxiety recognition model that the researchers created- low, mild, moderate and high, refer to the anxiety levels rather than distinct anxiety disorder classes. With signal fusion based SVM classifier, model achieved an accuracy of 80.1% using
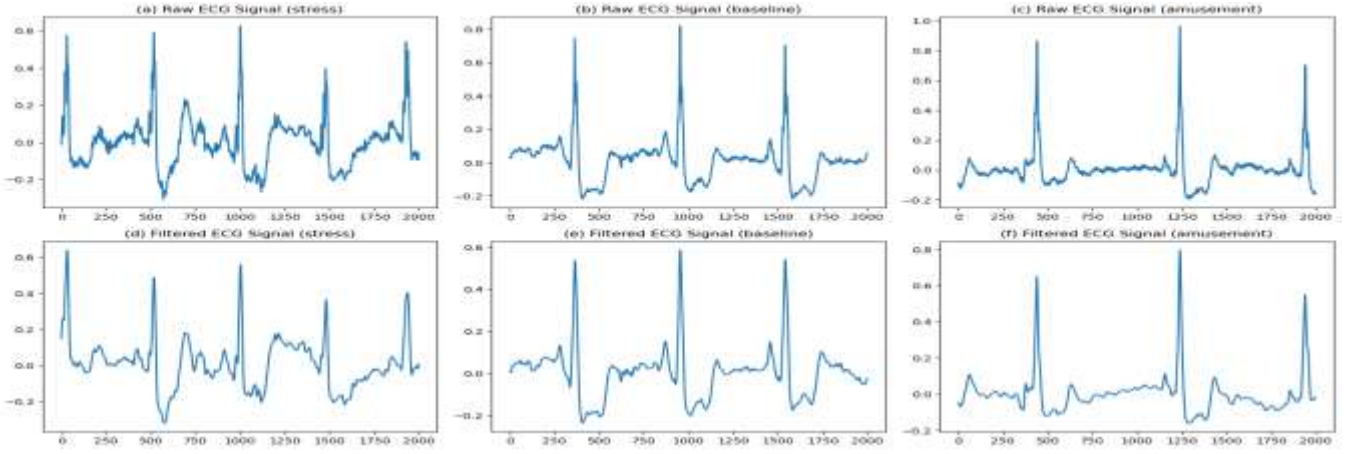
Fig 1. Raw vs Filtered ECG signal for stress, baseline and amusement condition

leave one subject out cross-validation and accuracy of 86.3% using 10×10 fold cross-validation.

A Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) based approach has also been proposed for emotion classification [9]. It was focused on novel trends in affective computing, leveraging reliable physiological signals like EEG, ECG, and GSR. The proposed approach involves a subject-independent computational framework, incorporating a 2D-CNN for EEG and a combination of LSTM and 1D-CNN for ECG and GSR. Two publicly available datasets, DREAMER [13] and AMIGOS [10], are used with low-cost, wearable sensors for real-world applicability. Results surpass state-of-the-art approaches, achieving high accuracy (98.73%) for emotion elicitation using ECG, and multi-modal fusion achieves the highest overall accuracy of 99.0% for the AMIGOS dataset and 90.8% for the DREAMER dataset.

III. DATASET DESCRIPTION

The developed approaches were validated against the WESAD dataset [4]. WESAD dataset was chosen over other datasets like the DREAMER [13] and CASE [16] because it contains multi-modal data that was captured in a variety of affective states, particularly stress. Schmidt et al.'s stress estimation procedure served as the basis for the collection of the WESAD dataset [4]. The workers' council and the research center's data security officer authorized the protocol. The dataset comprises mobility and physiological signals from 12 male and 3 female participants, with an average age of 27.5±2.4 years.

The dataset encompasses several sensor modalities such as blood volume pulse, electrocardiogram, electrodermal activity, electromyogram, respiration, body temperature, and three-axis acceleration. The data was recorded using two devices: a chest-worn device (RespiBAN) and a wrist-worn device (Empatica E4). The RespiBAN device provides data at a sampling rate of 700 Hz, while the Empatica E4 device provides data at varying sampling rates depending on the sensor. The data gathering procedure is briefly described below; further information is provided in [4]. Three conditions made up the majority of the experiment: baseline, stress, and amusement. After the stress and amusement conditions, the subjects were de-excited with a guided meditation. To maintain diverse postures in each condition, half of the individuals completed the experiment while sitting, and the other half performed it while standing. The subjects spent twenty minutes either standing or sitting while

reading neutral magazines during the baseline. The participants in the entertainment condition saw 11 humorous videos, spaced five seconds apart. This state lasted for around six minutes in total. The Trier Social Stress Test (TSST) was used to create tension during the stress situation [16]. Prior to taking the test, the participants had to prepare and deliver a five-minute personal speech in front of a three-person panel. Second, they had to count in increments of 17 from 2023 to zero, and in case they made a mistake, they had to start over. The stress situation lasted for almost ten minutes in total. The subjects were instructed to meditate for roughly seven minutes in order to return to their neutral condition. They were sitting comfortably and doing a breathing exercise while keeping their eyes closed.

Dataset contains the contextual features for each subject which includes information that are used in anxiety detection.In addition to the sensor and contextual data, the dataset also includes self-reports from the subjects, obtained using several established questionnaires. The assignment of anxiety classes in the study was based on a questionnaire called 6-STAI [5].

IV. METHODS

In this section, the raw signals were pre-processed and segmented into smaller segments. Following that, features were derived from the signals, and various data scaling and balancing techniques were applied before implementing machine learning methods. Once the features were derived from the segmented signals, the data underwent a series of processing steps to ensure optimal performance of the machine learning algorithms. Techniques such as data scaling and balancing were crucial in preparing the data for model training. By implementing these techniques, the accuracy and reliability of the machine learning methods were significantly enhanced, leading to more robust and efficient models for analysis and prediction.

A. Pre-processing of Raw Signals

A crucial step in obtaining accurate and reliable results is the pre-processing of raw physiological signals. Pre-processing involves signal extraction, filtering, noise removal, signal segmentation, and many more techniques. A person's state of emotions and cognition can be evaluated by using a variety of physiological signals, such as the heart rate (HR), GSR, ECG, and EEG [18]. These signals may be influenced by a number of factors, e.g. noise, movement, environment, etc. Preprocessing is therefore required to improve the signals' quality and lower their unpredictability.
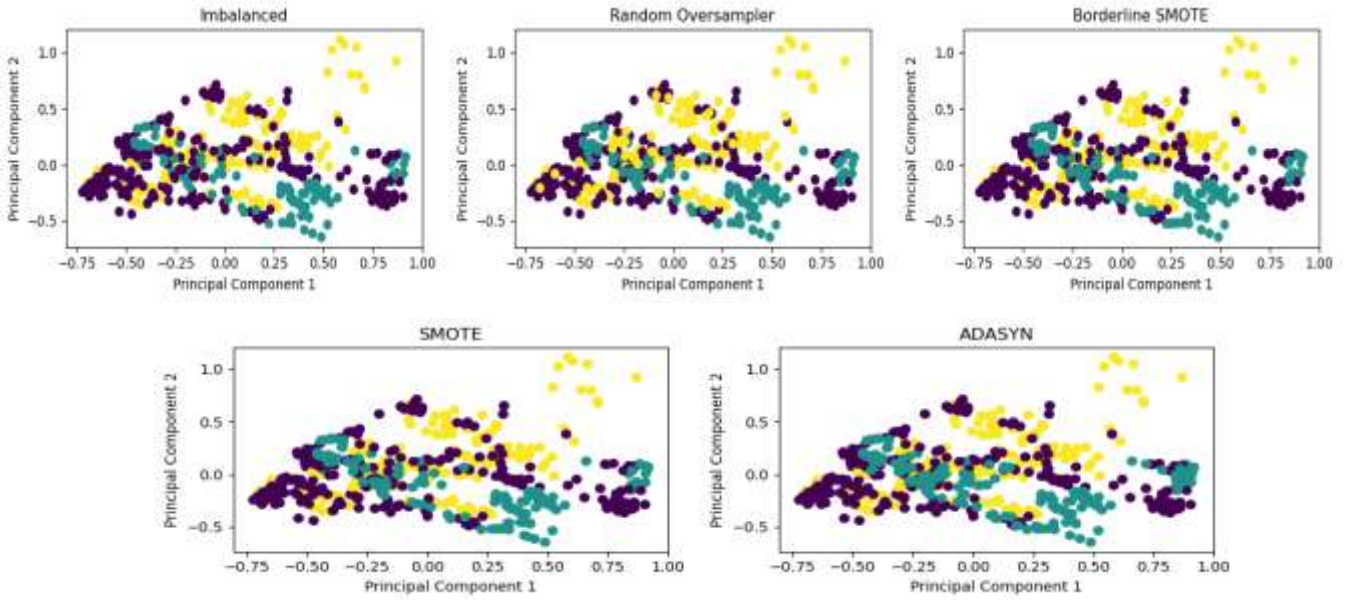
Fig 2. PCA analysis of actual imbalanced data and after balancing the dataset using random oversampled, borderline SMOTE, SMOTE and ADASYN algorithms.

ECG and EDA chest-data is extracted from the dataset. Signals under three conditions i.e. baseline, stress and amusement were extracted. ECG and EDA signals are segmented using a sliding window method with a sampling frequency of 700 Hz. with a window size of 60 seconds for every condition and every sample. Total number of samples generated was 535. After segmentation, raw signal filtering is done. ECG data has high frequency noises which includes electromyogram noise and additive white Gaussian noise and it also has low frequency noises which includes baseline wandering (movement and respiration) [11]. A Butterworth bandpass filter of order 4 is applied between 0.5-40 Hz. Fig.1 shows raw and filtered ECG signal segments for a subject under all the three conditions. EDA data contains electrical interference noise, so a 2nd-order butterworth lowpass filter is applied.

B. Feature Extraction and Scaling

Feature extraction is a critical component in learning a model from physiological information. After pre-processing, the filtered segments are ready for the feature extraction procedure. Features that are extracted from ECG and EDA signals are shown in Table 1. A total of 16 features are extracted from these physiological signals and used for classification of anxiety.

a) *ECG feature extraction*: From filtered ECG signals, statistical features such as mean, variance,standard deviation, minimum, maximum are calculated and peak detection algorithms are applied to detect the peaks from which heart rate(HR) and RR-intervals are calculated. RR-interval is the time between successive R peaks and provides information about heart rate variability (HRV). Time domain measures such as mean RR-intervals, standard deviation of RR-intervals (SDNN), triangular interpolation (TINN) and root mean of sum of successive differences (RMSSD) are calculated. Reduced SDNN and RMSSD have been associated with increased anxiety [19].

b) *EDA feature extraction:* EDA signal is made up of two components: phasic component (SCR) and tonic component (SCL). Phasic and tonic components are separated using physiology package [12]. Mean peak duration and mean peak magnitude for SCR is computed.

Statistical features like mean, standard deviation, minimum and maximum of EDA signals are also calculated.

TABLE I. LIST OF EXTRACTED FEATURES

Signal	Features Extracted
ECG	Min, Max, mean, variance ,standard deviation of ECG, Heart rate, meanNN, RMSSD, SDNN, TINN
EDA	Min, Max, mean, standard deviation of EDA, Mean peak duration of SCR, Mean peak magnitude of SCR

Because outliers may indicate more severe cases of anxiety, they are included in this research because they may represent significant cases. In order to prevent features with higher values from predominating, features are scaled. It's crucial to stay away from subject-wise dependency as well. Min-Max scaling is used here for normalization of points and mapping the data to [0,1] intervals.

C. Context Features

Contextual features are features that capture the information about the situation or environment in which the physiological signals are recorded. Contextual features can help to improve the accuracy and reliability of anxiety detection models by accounting for the variability and noise in the physiological data. Contextual features can also provide insights into factors that can influence anxiety levels of the individuals. The phases of the experimental protocol, such as baseline, stress induction, and recovery [21]. These phases can indicate the expected level of anxiety and the temporal dynamics of the physiological responses. The type and intensity of the stressor, such as social, physical, or cognitive. Different stressors can elicit different patterns of anxiety and physiological reactions. The demographic and personal characteristics of the individuals, such as age, gender, personality, and mood [20]. These factors can affect the susceptibility and resilience to anxiety and the expression of physiological signals. The contextual features for each subject which includes, whether he/she has taken coffee today or within the last hour of the experiment, whether he/she has done any sports today, whether he/she smokes or has smoked within the last hour, whether he/she feels ill and height and

weight from which BMI is calculated. These context features are used to create context feature vectors.

D. Self Report Analysis

The 6-STAI questionnaire score was calculated for each individual during the baseline, stress and amusement condition and has been shown in Fig 3. The questionnaire contained six items, namely, Feel at ease, nervous, jittery, relaxed, worried and pleasant.Each item was scored between 1 to 4 by every subject after each condition. The total scores have been calculated and categorized into three classes: low(1-11.1), moderate(11.4-13.2) and high(13.4-24) [5]. By analyzing the self reports, it is found that the anxiety score distribution is not balanced and can be seen in Fig.3(a). Participants with "Moderate" anxiety levels are very few as compared to "Low" and "High" anxiety levels.
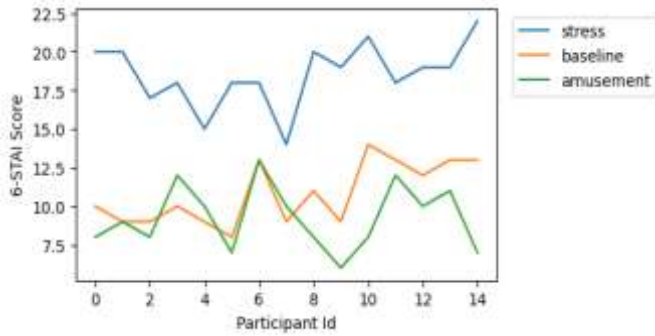

Fig 3. Barplot showing number of participants in each class

E. Data Balancing Techniques

Dataset is imbalanced, with 243, 123 and 169 samples of low, moderate and high anxiety levels respectively. There are several data balancing techniques. Oversampling techniques are applied as the number of samples is only 535. Some of the data balancing algorithms are applied like Random Oversampling, SMOTE, Borderline SMOTE and ADASYN. Randomly adding instances from the minority class to the training dataset is known as random oversampling. In order to enhance random oversampling, SMOTE was created [14]. It functions by selecting a random minority class sample, determining the example's k nearest neighbors, and then interpolating the feature values of the two instances to create a new synthetic sample. Borderline SMOTE is an addition in the SMOTE algorithm. Borderline SMOTE focuses specifically on the instances that are near the decision boundary (borderline instances). ADASYN computes a density distribution of minority instances and generates more synthetic samples for instances in regions where the class imbalance is higher. This adaptive approach helps in addressing the imbalance by providing more emphasis on the instances that are harder to learn [17]. As shown in Fig.2, there is a slight increase in the number of minority samples however with some overlapping. Results will be analyzed with each of these algorithms.

F. Classification

Classification strategies have been employed to categorize the anxiety into three classes: low, moderate, and high. Table 2 displays a comparison of the anxiety class classification results with and without contextual features. Results from Decision Tree and Gradient Tree Boosting with single and multimodal physiological signals are also compared. Leave-

one-out cross validation (LOOCV) was used to assess the machine learning model's performance.

V. RESULTS

This section presents the results of the classification of anxiety into three classes, namely low, moderate and high with single and multimodal physiological signals. A comparison between accuracies of Decision Tree (DT) and Gradient Tree Boosting (GBDT) with different data balancing techniques for single modal and multimodal physiological signals with and without context features is shown in Table 2. Performance metric that is used here is accuracy. In single modal classification, taking ECG into consideration, Random Oversampling with Gradient Tree Boosting and LOOCV gave best results with 89.8% accuracy. With EDA signals, Random Oversampling with Gradient Tree Boosting gave the highest accuracy of 86.3%. Results of multimodal classification are better than the single modal classification as can be seen in Fig 4(a), where Gradient Tree Boosting algorithm with LOOCV outperformed when implemented with every mentioned oversampling technique with highest accuracy of 96.7% with Random Oversampling when calculated without taking context features into consideration. Multimodal physiological signal classification with context features gave the best results with accuracy of 97.3% using GBDT and LOOCV. A confusion matrix for multimodal classification with context features using Gradient Tree Boosting algorithm is shown in Fig 5.
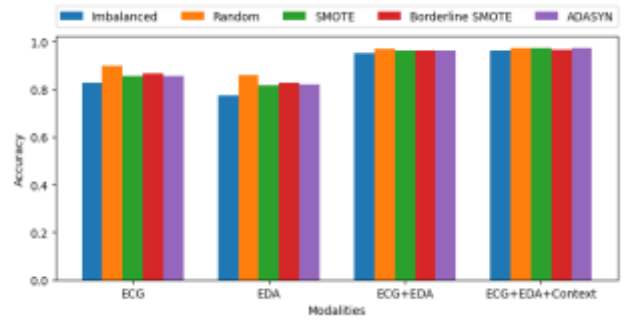

Fig 4. Anxiety classification results for different modalities with different data balancing techniques

VI. CONCLUSION

Research was performed on the WESAD [4] dataset and results were calculated using single modalities as well as multi modalities with context features. Physiological signals that are used here are, EDA and ECG collected under baseline, stress and amusement conditions. Pre-processing of signals is done by applying various filters and features are extracted from these filtered signals. After applying Decision Tree Classifier and Gradient Tree Boosting Classifier on the dataset, It was discovered that using context features with multimodal physiological signals gave better results as compared to classification without context features. Context features provided insights about the individuals' which contributed in the classification of anxiety levels. Only two physiological signals were taken under consideration, namely ECG and EDA.To get more accurate results, more physiological signals can be fused together to see more accurate results. Also, another limitation is the less number of samples in the dataset, which led to less representativeness in the classification.
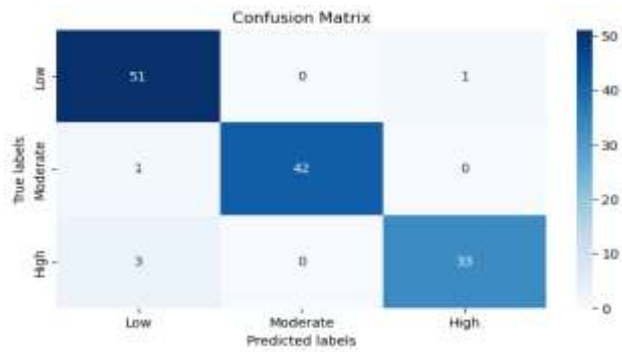
Fig 5. Confusion matrix of Gradient Tree Boosting algorithm with ECG, EDA and context features.

TABLE II. COMPARISON BETWEEN RESULTS OF MACHINE LEARNING MODELS WITH DIFFERENT BALANCING TECHNIQUES

Accuracy						
Modalities	Algorithm	Imbalanced	Random Oversampling	SMOTE	Borderline SMOTE	ADASYN
ECG	DT	75.7%	89%	83.2%	85.4%	77.4%
	DT+LOOCV	74.2%	88.5%	84.7%	87.3%	83.1%
	GBDT	79.4%	87.6%	84.7%	83.2%	80.4%
	GBDT+LOOCV	82.8%	89.8%	86.6%	87.3%	86%
EDA	DT	71%	82.1%	70.9%	73.2%	74.8%
	DT+LOOCV	72.9%	84.1%	75.4%	76.2%	73.1%
	GBDT	72.8%	86.3%	73.2%	77%	83.2%
	GBDT+LOOCV	77.2%	85.9%	82.3%	81.7%	82.4%
ECG + EDA	DT	81.3%	89%	84.7%	84.7%	84.7%
	DT+LOOCV	88.4%	90%	91.1%	89.9%	91.2%
	GBDT	91.5%	96.5%	92.3%	91.6%	91.6%
	GBDT+LOOCV	95.1%	96.7%	96.2%	96.6%	95.5%
ECG + EDA + Context	DT	85.9%	87.6%	87.7%	88.5%	93.9%
	DT+LOOCV	89.9%	91.8%	91.6%	93.6%	92.6%
	GBDT	91.5%	93.1%	95.4%	96.9%	93.1%
	GBDT+LOOCV	96.3%	97.3%	97.1%	96.5%	96.7%

REFERENCES

[1]. D. F. Santomauro et al., "Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic," The Lancet, vol. 398, no. 10312, pp. 1700-1712, 2021.

[2]. S. S. Panicker and P. Gayathri, "A survey of machine learning techniques in physiology based mental stress detection systems," Biocybernetics and Biomedical Engineering, vol. 39, no. 2, pp. 444-469, 2019.

[3]. N. Sharma and T. Gedeon, "Objective measures, sensors and computational techniques for stress recognition and classification: A survey," Computer methods and programs in biomedicine, vol. 108, no. 3, pp. 1287-1301, 2012.

[4]. P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, "Introducing wesad, a multimodal dataset for wearable stress and affect detection," in Proceedings of the 20th ACM international conference on multimodal interaction, 2018, pp. 400-408.

[5]. T. M. Marteau and H. Bekker, "The development of a six-item short-form of the state scale of the Spielberger State—Trait Anxiety Inventory (STAI)," British journal of clinical Psychology, vol. 31, no. 3, pp. 301-306, 1992.

[6]. M. Adheena, N. Sindhu, and S. Jerritta, "Physiological Detection of Anxiety," in 2018 International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET), 2018: IEEE, pp. 1-5.

[7]. M. Vaz, T. Summavielle, R. Sebastião, and R. P. Ribeiro, "Multimodal Classification of Anxiety Based on Physiological Signals," Applied Sciences, vol. 13, no. 11, p. 6368, 2023.

[8]. J. Šalkevicius, R. Damaševičius, R. Maskeliunas, and I. Laukienė, "Anxiety level recognition for virtual reality therapy system using physiological signals," Electronics, vol. 8, no. 9, p. 1039, 2019.

[9]. M. N. Dar, M. U. Akram, S. G. Khawaja, and A. N. Pujari, "CNN and LSTM-based emotion charting using physiological signals," Sensors, vol. 20, no. 16, p. 4551, 2020.

[10]. J. A. Miranda-Correa, M. K. Abadi, N. Sebe, and I. Patras, "Amigos: A dataset for affect, personality and mood research on individuals and groups," IEEE Transactions on Affective Computing, vol. 12, no. 2, pp. 479-493, 2018.

[11]. L. Sörnmo and P. Laguna, Bioelectrical signal processing in cardiac and neurological applications. Academic press, 2005.

[12]. Gabrieli, A. Azhari, and G. Esposito, "PySiology: A python package for physiological feature extraction," Neural approaches to dynamics of signal exchanges, pp. 395-402, 2020.

[13]. S. Katsigiannis and N. Ramzan, "DREAMER: A database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices," IEEE journal of biomedical and health informatics, vol. 22, no. 1, pp. 98-107, 2017.

[14]. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," Journal of artificial intelligence research, vol. 16, pp. 321-357, 2002.

[15]. K. Sharma, C. Castellini, E. L. van den Broek, A. Albu-Schaeffer, and F. Schwenker, "A dataset of continuous affect annotations and physiological signals for emotion analysis," Scientific data, vol. 6, no. 1, p. 196, 2019.

[16]. C. Kirschbaum, K.-M. Pirke, and D. H. Hellhammer, "The 'Trier Social Stress Test'–a tool for investigating psychobiological stress responses in a laboratory setting," Neuropsychobiology, vol. 28, no. 1-2, pp. 76-81, 1993.

[17]. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in 2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence), 2008: Ieee, pp. 1322-1328

[18]. V. Khullar, R. G. Tiwari, A. K. Agarwal, and S. Dutta, "Physiological signals based anxiety detection using ensemble machine learning," in Cyber Intelligence and Information Retrieval: Proceedings of CIIR 2021, 2022: Springer, pp. 597-608.

[19]. R. Hartmann, F. M. Schmidt, C. Sander, and U. Hegerl, "Heart rate variability as indicator of clinical state in depression," Frontiers in psychiatry, vol. 9, p. 735, 2019.

[20]. E. Calikus, S. Nowaczyk, M.-R. Bouguelia, and O. Dikmen, "Wisdom of the contexts: active ensemble learning for contextual anomaly detection," Data Mining and Knowledge Discovery, vol. 36, no. 6, pp. 2410-2458, 2022.

[21]. R. K. Nath and H. Thapliyal, "Machine learning-based anxiety detection in older adults using wristband sensors and context feature," SN Computer Science, vol. 2, no. 5, p. 359, 2021.