IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 9 August 2024, accepted 10 September 2024, date of publication 16 September 2024, date of current version 22 October 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3461588

RESEARCH ARTICLE

Resp-BoostNet: Mental Stress Detection From Biomarkers Measurable by Smartwatches Using Boosting Neural Network Technique

SANJAY KUMAR[1], ANSHUMAN RAJ CHAUHAN[1], AKHIL[1], AKSHI KUMAR[2], (Senior Member, IEEE), AND GUANG YANG[3,4,5,6], (Senior Member, IEEE)

[1]Department of Computer Science and Engineering, Delhi Technological University, New Delhi 110042, India
[2]Department of Computing, Goldsmiths, University of London, WC1E 7HU London, U.K.
[3]Bioengineering Department and Imperial-X, Imperial College London, W12 7SL London, U.K.
[4]National Heart and Lung Institute, Imperial College London, SW7 2AZ London, U.K.
[5]Cardiovascular Research Centre, Royal Brompton Hospital, SW3 6NP London, U.K.
[6]School of Biomedical Engineering and Imaging Sciences, King's College London, WC2R 2LS London, U.K.

Corresponding author: Guang Yang (g.yang@imperial.ac.uk)

This work was supported in part by ERC IMI under Grant 101005122, in part by H2020 under Grant 952172, in part by MRC under Grant MC/PC/21013, in part by the Royal Society under Grant IEC\NSFC\211235, in part by the NVIDIA Academic Hardware Grant Program through the SABER Project supported by Boehringer Ingelheim Ltd., in part by Wellcome Leap Dynamic Resilience through UKRI guarantee funding for Horizon Europe MSCA Postdoctoral Fellowships under Grant EP/Z002206/1, and in part by the UKRI Future Leaders Fellowship under Grant MR/V023799/1.

ABSTRACT To maintain overall health and well-being, it is crucial to manage mental stress. This study focuses on developing a deep learning model for recognizing mental stress levels using the sensors of smartwatches. Most related research with notable performance has focused on mental stress detection using various physiological biomarkers obtained through sophisticated IoMT (Internet of Medical Things) devices. However, the ones using only the smartwatch's measurable physiological biomarkers, which do not include respiration rate, have comparatively lower performance because of a limited number of physiological biomarkers. In this paper, we introduce an improved model for mental stress detection using boosting neural network that can be integrated into a smartwatch. The proposed model consists of two phases. In the first phase, we introduce a boosting neural network technique that predicts the respiration rate by utilizing the biomarkers measurable by a smartwatch. In the second phase, the modified set of biomarkers, which includes both the original biomarkers and the predicted respiration rate, is used for stress level classification via an artificial neural network. The necessary hyperparameter tuning is performed to obtain the optimal values of various model parameters. The training of the model is performed for fifteen different subjects of the publicly available multimodal WESAD (Wearable Stress and Affect Detection) dataset using various biomarkers measured by smartwatches. The proposed model predicts respiration rate with low error (0.035 MSE (Mean Squared Error)) and achieves high mental stress detection accuracy of 94% using smartwatch measurable biomarkers which is a ~2% improvement over the current contemporary technique.

INDEX TERMS Artificial neural network, boosting neural network, deep learning, mental stress detection, regression, respiration rate.

I. INTRODUCTION

Mental stress can be termed as psychological or emotional strain triggered by physical or mental discomfort felt while

The associate editor coordinating the review of this manuscript and approving it for publication was Binit Lukose.

performing some demanding task or in some cases while performing some daily life routine. Mental stress can be caused by various factors such as work pressure, financial difficulties, relationship issues, traumatic events, or chronic health conditions. Mental stress is not necessarily a negative thing and can be considered a normal component of life. It can be classified as positive mental stress which is generally acute mental stress and negative mental stress which is chronic mental stress or long-term mental stress [1]. In certain situations, to get optimal human output or to enter a state of motivation, short-term stress plays a vital role. But prolonged mental stress results in various health issues such as slower body recovery, raised blood pressure, bad sleep quality, increased vulnerability to infections, and decreased mental performance [2]. Prolonged or frequent exposure to stressors, which can be the result of performing vigorous activities, can lead to chronic mental stress. While performing vigorous activities the body undergoes a noticeable physical and psychological change that results in increased metabolic demand [3], [4]. To fulfill the increased demands, the body's sympathetic nervous system speeds up certain processes within the body such as heart rate, respiration rate, sweating rate, etc. The sympathetic nervous system slows down and the parasympathetic nervous system begins the rest and repair process [5] after the completion of intense activity. A balance between the two processes is desired for the normal functioning of the body. However, frequently working in a stressful environment can lead to higher responses from a sympathetic nervous system that may result in chronic mental stress. Chronic mental stress has an adverse effect on not just the mental but also the physical health of a person. Citing the various ill effects of mental stress and the increased mental stress levels across the population throughout the world due to the recent pandemic situation [6], a mental stress detection system has become a must for the self-management of mental stress.

There is a lack of awareness among people regarding mental stress which has made people unaware of their mental stress behavior. Due to this reason, it becomes difficult to collect stress-related data for research purposes. Often time researchers have to rely on self-assessments provided by volunteers [7]. Since stress condition is accompanied by some physiological and psychological changes, the usage of sensors to measure those changes provides another means to collect the data [8]. Surveys based on self-assessments often time provides biased data which makes the data quite unreliable. So, instead of relying on surveys for data collection, more accurate and reliable data can be collected easily using modern-day's advanced sensors that make measurements along three main modalities namely physiological changes, psychological changes, and behavioral changes. Mental stress detectors that can capture specific and effective physiological effects of mental stress are quite reliable in terms of physiological data collection. Physiological data is robust to distortions and it also does not get obscured. *Biopacs, RespiBAN professional, Empatica, E4, MP150* etc can be used to measure physiological changes to produce good results. Few researchers have tried creating datasets for mental stress detection using these devices by measuring various physiological data. Healey and Picard [9] published a dataset with features such as ECG (Electrocardiogram), EDA (Electrodermal Activity), RESP (Respiration Rate), and EMG (Electromyogram) data on driver stress. Koestra et al. [10] published DEAP (Dataset for Emotion Analysis using EEG, Physiological and video signal), a database for emotion analysis using physiological signals. The dataset contained EEG (Electroencephalogram) signals, facial videos, and peripheral physiological signals. More recently, Schmidt et. al. [11] published WESAD. This dataset has a rich set of features that includes signals such as EMG, EDA, ECG, RESP, Body Temperature, and three axes Acceleration. However, using sensor-based devices has its own limitations because of the subjective nature of mental stress. Different people perceive mental stress in different ways and their response (physiological changes) to mental stress also varies. Hence citing the reliability factor, a better alternative becomes combining sensor measurements with self-assessments. Wrist-worn devices can recognize mental stress quite reliably and many recent studies have been focussed on improving the reliability of predictions of these devices [12]. For instance, Gjoreski et. al. [13] used skin temperature, electrodermal activity, heart rate, blood volume pulse, and accelerometer signals to detect mental stress levels under laboratory conditions and were able to detect it with an accuracy of 83%. Remirez et al. [14] also used physiological biomarkers (heart rate, galvanic skin response and breath) for mental stress detection.On the other hand, Bannore et. al. [15] used facial expression recognition system to detect stress levels. Many of the studies related to mental stress detection use respiration rate in the recognition process [16] along with other general biosensors.

Many of the existing studies to detect mental stress use sophisticated chest-wearable IoMT devices for the measurement of physiological biomarkers [11]. However, wearing such devices is impractical for daily use. On the other hand, the studies utilizing physiological biomarkers collected through convenient devices such as a smartwatch, exhibit comparatively lower performance due to a limited range of physiological indicators. One such missing physiological biomarker is respiration which is an important indicator of mental stress as evidenced by multiple research showing a positive correlation between mental stress and respiration rate [17], [18], [19]. Our findings from a series of experiments also aligned with these researches and we also conclude that respiration rate is an important biomarker that can be used to detect mental stress levels with high accuracy (Tab. 5).

In this study, we intend to propose a method that utilizes a convenient wearable device (a smartwatch) for detecting mental stress with high accuracy. The proposed model for mental stress detection, employs a BNN (Boosting Neural Network) to predict respiration rate using smartwatch measurable physiological biomarker and combines it with the biomarker set to ultimately predict stress levels using an ANN (Artificial Neural Network). With the help of some compression techniques [20], [21], [22], [23] the proposed model can be easily integrated within smartwatches. The proposed model consists of two phases; the first one is a regression task for predicting respiration rate and the second one is a classification task for detecting the current mental stress level (affective state). Combining the predicted respiration rate as a feature in the existing data resulted in performance improvement of the classifier as compared to its performance in the absence of a respiration rate. We compare the performance of the proposed model with various baseline and contemporary methods of mental stress detection. The obtained results reveal the superiority of the proposed model over other comparing models. The major contributions of this work are as follows.

1) A deep learning-based improved model, named Resp-BoostNet, is introduced for detecting mental stress from physiological biomarkers measurable by smartwatches.
2) We effectively perform dataset aggregation to reduce the number of samples and dataset balancing, without any qualitative loss of data.
3) A Boosting Neural Network-based regression model is introduced to predict respiration rate using the physiological biomarker measured by a smartwatch and append that data to the feature vector of each sample from the dataset.
4) We perform intensive experimentation using the WESAD dataset and evaluate various performance metrics. The obtained results reveal the superior performance of the proposed model against the recently proposed model for mental stress detection.

The organization of the rest of the paper is as follows: Section II comprises the related work. A brief description of the dataset and evaluation metrics used is provided in Section III. Section IV contains details about the proposed methodology. Experimental analysis and performance evaluation are illustrated in Section V followed by concluding remarks and future scope in Section VI.

II. RELATED WORK

Due to the consequences of mental stress on health, there is extensive research on mental stress recognition. This section discusses some of the recent work in the domain of mental stress detection.

The frequency at which multi-modal devices take measurements is generally quite high which results in the generation of tons of data. Thus, the focus of some research pivots around finding efficient data clustering techniques to reduce the number of data points to make it easier for stress detection models. In this direction, Kumar et al. [24] proposed a genetically optimized Fuzzy C-means data clustering technique. The clustering helped in summarizing the data for each subject. They used a multi-level CNN (Convolutional Neural Network) architecture for the classification task. Gupta et al. [25] proposed a similar method for clustering. They utilized time-frame restricted intra-similarity calculations to perform k-medoid data clustering to obtain summarized data. Hovsepian et al.'s [26] work primarily focused on the efficient pre-processing of data recorded by an AutoSense sensor. They performed interpolation and normalization on the ECG and respiration rate data and then used the refined data for mental stress prediction. They achieved an accuracy of 89% on the validation set and 72% on the test set using an SVM (Support Vector Machine) classifier. Sharma et al. [27] extracted entropy based features from EEG signals to detect the mental stress level using a support vector classifier. To enhance the performance of their SVM classifier they also used evolutionary optimizing techniques. Samarpita and Satpathy [28] also worked with EEG signals and proposed an algorithm for stress level detection using EEG data. They used the bandpass filter method to separate EEG signals into EEG rhythms and normalized EEG signals using the k-mean clustering method.

Most of the research in the domain of mental stress detection combines the multi-modal data collected from wearable smart devices and sensors with machine learning techniques to estimate human affective state [29], [30], [31], [32], [33], [34]. Sagbas et al. [35] and Garg et al. [36] leveraged smartphone and wearable devices for mental stress detection while performing daily routine activities. Sagbas et al. [35] also used smartphones to collect accelerometer and gyroscope sensor data of writing behavior for 46 subjects. They used several machine learning algorithms for classification and obtained the highest accuracy of 87.56% for the k-nearest Neighbour method. Gjoreski et al. [30] devised a method for continuous mental stress detection using data from wrist devices. Their method had three components: a laboratory mental stress detector, built using a random forest classifier, for detecting short-term mental stress; an activity recognizer to continuously monitor the user's activity; and a context-based mental stress detector that leverages the outputs of other two components to generate the final decision after every 20 minutes. This component was developed using an SVM classifier. Siirtola [12] demonstrated the mental stress detection accuracy for various standard classifiers using smartwatch measurable features. The author also studied the importance of various physiological biomarkers for mental stress detection and concluded that EDA signals are unnecessary for recognizing mental stress when the window length is big enough. The window length here refers to the length of time considered for converting analog signals such as ECG to discrete data points. Zhu et al. [37] examined the classification performance of four different electrodermal activity databases for stress detection using several machine learning models. They achieved an accuracy of 92.9 for the approach using SVM.

Some recent researchers have tried to leverage the power of different deep-learning techniques for mental stress detection. Kumar et al. [38] proposed a multi-level deep neural network having hierarchical learning capabilities of CNN. To combine the high-level features into a single unified representation they used a model-level fusion strategy. The proposed model had an accuracy of 87.7% on the WESAD dataset. Yu et al. [39] proposed modality fusion and personalized attention mechanism for dealing with missing modalities and adapting the classification model for every user differently. Something along a similar line was done by Ahmed et al. [40]. They proposed a method that develops a multimodal diagnosis system that takes patient specific factors into consideration. Their model consists of an attention based multimodal classifier with selective dropout and normalization. Through rigorous experimentation on three different multimodal datasets, they achieved an F1 score of 0.945. Khan [41] performed a comparison between supervised that used CNN combined with bi-directional LSTM (Long Short Term Memory) and semi-supervised (used Generative Adversarial Network) techniques to detect mental stress. Their study showed that semi-supervised techniques fairly outperformed fully supervised techniques.

Mou et al. [50] used an attention based CNN and LSTM model for driver stress detection. Iqbal et al. [51] tried to assess the relative sensitivity and specificity of common physiological biomarkers that are indicators of mental stress. They performed a statistical analysis of various physiological data gathered from healthy individuals and concluded that respiration rate and heart rate are the two best features for detecting the state of mental stress. Indikawati and Winiarti [52] used standard machine learning classifiers to develop a personalized mental stress detection system using the WESAD dataset. They were able to achieve quite high accuracy for specific subjects. In summary, Tab. 1 lists the various research works studied in this article. The papers are listed in the order of appearance in the related work section.

All these works either used a limited number of physiological biomarkers, which limits their performance, or used physiological biomarkers that are measured by some sophisticated devices that are not suitable to be worn on a daily basis. This creates a trade-off between the performance of the stress detection model and the convenience of the user. The proposed work intends to bridge this gap by generating an additional physiological biomarker with the help of a strong deep learning based regressor, using only the smartwatch's measurable physiological biomarkers.

III. DATASETS AND EVALUATION METRICS

This section describes the details of the dataset used for training the proposed model and also the various evaluation metrics used to evaluate the performance of the proposed model.

A. DATASET

The model has been trained on the publicly available WESAD dataset. Schmidt et. al. [11] performed a lab study to collect the physiological and motion data of 17 subjects. During the experiments the devices attached to two of the subjects, for measuring their physiological changes, malfunctioned and hence the dataset contains data for only 15 subjects. The measurements were taken for a period of two hours through chest wearable (*RespiBAN Professional*) and wrist wearable (*Empatica E4*) IoMT device. A total of 12 biomarkers were measured out of which we used six in this work. The details of the same are listed in Tab. 2.

The dataset contains three different affective states (neutral, stress, and amusement) which is an improvement over previous datasets in this field which were majorly limited to two affective states, namely stress and no stress.

B. EVALUATION METRICS

Evaluation metrics are used to assess the performance of an algorithm. For assessing the performance of our proposed work we used various evaluation metrics. The usage of multiple evaluation metrics helps in better assessment of the performance of different techniques. For assessing the performance of classification tasks we used accuracy, precision, recall, and F1 score, while for regression tasks we used mean square error.

IV. PROPOSED WORK

In this section, we present our proposed model for mental stress detection using various physiological biomarkers measured by a smartwatch. First, we perform necessary data pre-processing, which consists of four tasks; (i) Data discretization to generate discrete samples from time series data, (ii) Data aggregation to reduce the number of samples for computational feasibility, (iii) Data set balancing to minimize chances of biasing and (iv) Data normalization for improving model's performance. The proposed model consists of two phases: (a) Regression task for predicting respiration rate for which we combined boosting technique with a neural network and (b) Classification task for detecting mental stress levels using an ANN.

Using the pre-processed data, we tried to ascertain the importance of each biomarker. For this task, we remove each biomarker, one at a time, from the feature vector of the samples of the dataset while keeping the rest biomarkers in the feature set and perform the classification task using an ANN. This analysis helped in assessing the effectiveness of all the biomarkers in detecting mental stress levels. We found that respiration rate, a biomarker not measured by currently available smartwatches (although in the WESAD dataset, ECG and EMG are measured by a chest-worn device, they can be measured by currently available smartwatches [53]), had a significant role in deciding the affective state of a person. Our analysis pointed out that the inclusion of

TABLE 1. A summary of various research papers studied by us in the related work section.

Ref No. & Year	Dataset used	Techniques used	Result
[12] 2019	WESAD	1. The importance of various physiological biomarkers was analyzed for mental stress detection 2. The accuracy of various standard machine learning classifiers for mental stress detection was demonstrated.	1. An accuracy of 87.4% was achieved. 2. Authors concluded that for larger window length (time interval used for discretization of analog signals) EDA signals are unnecessary for mental stress detection.
[24] 2021	WESAD	1. A genetically optimized Fuzzy C-means data clustering technique was used to generate summarized data. 2. A multi-level CNN was used for mental stress prediction	An average model accuracy of 87.7% was achieved.
[25] 2021	WESAD	1. Time-frame restricted intra-similarity calculations were used to perform k-medoid data clustering to obtain summarized data. 2. A CNN was trained on the summarized dataset for classification.	A reduction of 34% in the average execution time was achieved with comparable accuracy to the original WESAD paper.
[26] 2015	Collected data from an independent lab study with 21 participants	1. Interpolation and normalization on the ECG and respiration rate data were performed and then used the refined data for mental stress prediction using SVM classifier.	An accuracy of 72/% was achieved.
[27] 2022	EEG during mental arithmetic tasks dataset [42]	1. EEG signals were decomposed using stationary wavelet transform. 2. Entropy-based features were extracted from the decomposed signals. 3. SVM classifier was used for the classification of affective states. To improve the performance of a classifier, its parameters were optimized using different evolutionary-inspired approaches.	SVM optimized using whale optimization algorithm resulted in the best accuracy of 97.3%.
[30] 2017	Collected data from an independent lab study with 21 participants	1. This method had three components: - A laboratory mental stress detector, for detecting short-term mental stress. - An activity recognizer to continuously monitor the user's activity. - A context-based mental stress detector that leverages the outputs of other two components to generate the final prediction.	The method had a recall of 70% and a precision of 95%.
[37] 2023	CLAS [43], UTD [44], VerBIO [45], and WESAD	1. Different machine learning models were used for stress detection using four different electrodermal activity databases.	1. The SVM model outperformed other techniques and achieved an accuracy of 92.9% for binary classification. 2. The authors concluded that EDA signals significantly improve the model performance.
[38] 2021	WESAD	1. The authors combined the high-level features of data into a single unified representation using a model-level fusion strategy. 2. A multi-level deep neural network having hierarchical learning capabilities of CNN was used for classification.	A performance accuracy of 87.7% was achieved.
[39] 2021	Collected wearable sensor and self-report data from 41 participants	1. A modality fusion and personalized attention mechanism were proposed for dealing with missing modalities and adapting the classification model for every user individually.	The personalized attention strategy had a 77.4% f1-score.
[40] 2023	D-vlog [46], DAIC-WOZ [47] [48], MODMA [49]	1. This method uses a multimodal diagnosis system that takes patient specific factors into consideration. 2. The classifier model consists of an attention based multimodal classifier.	Through rigorous experimentation on three different multimodal datasets, they achieved an F1 score of 0.945.
[41] 2022	WESAD	1. The authors majorly performed a comparative study between supervised (used CNN combined with bi-directional LSTM) and semi-supervised (used Generative Adversarial Network) technique's performance for mental stress detection.	Their study showed that semi-supervised techniques fairly outperformed fully supervised techniques.
[50] 2021	Collected data using a driving simulator, sampled for 22 participants from multiple driving scenarios.	1. The authors used an attention based CNN and LSTM model to fuse non-invasive data that include data collected from driving simulator such as eye data, vehicle data, and environmental data. 2. Usage of self-attention mechanism in their work allows assigning different levels of attention to features from different modalities.	The authors achieved an accuracy of 95.5% for driver stress detection.

respiration rate boosted the performance of the classifier. Hence, we propose to use a regression model to predict the data for respiration rate using other physiological biomarkers measured by smartwatches, append the predicted respiration

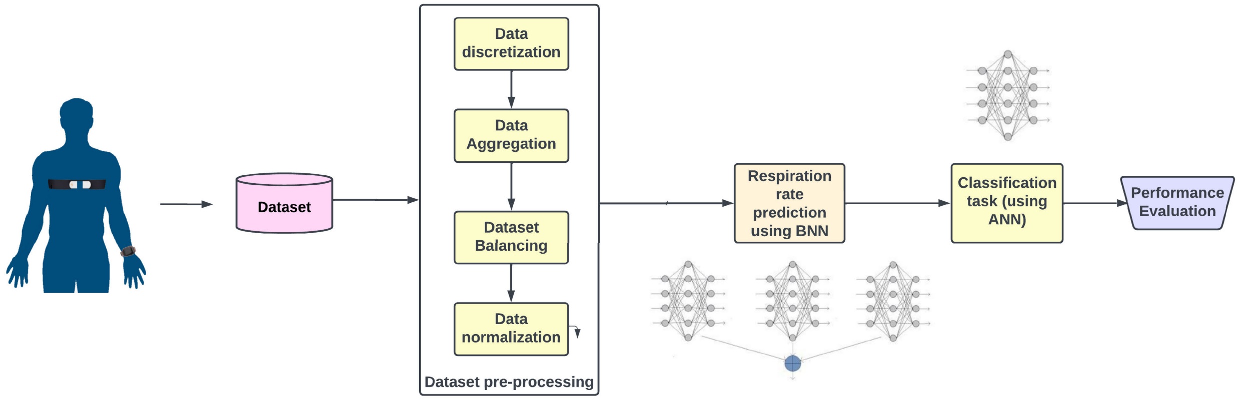


FIGURE 1. Flow diagram of the proposed boosting neural network based framework for stress detection using wearable smartwatches.

TABLE 2. Some of the physiological biomarkers present in the WESAD dataset.

Biomarker	Physiological changes measured	Measuring device
Electrocardiogram	Frequency of cardiac cycle	RespiBAN
Electromyogram	Musculoskeletal movements Helps in recognizing hand and face gestures.	RespiBAN
Body Temperature	Skin temperature	RespiBAN, Empatica E4
Respiration	Rate of breathing and exhalation	RespiBAN
Electrodermal Activity	Skin conduction	RespiBAN, Empatica E4
Three Axis Acceleration	Motion change in three dimensions	RespiBAN, Empatica E4

TABLE 3. Table of description.

Symbol	Description
X	Set of datapoints
X_i	i^{th} datapoint in the set of datapoints X
$X_{maximum}$	Maximum value in the set of datapoints X
$X_{minimum}$	Minimum value in the set of datapoints X
X_{in}	Normalised value of X_i
D	Dataset
n	Number of samples in dataset
d	Dimension (number) of feature in dataset
x_i	Feature set of i^{th} datapoint of dataset
y_i	Label of i^{th} datapoint of dataset
$\mathbb{R}$	Set of real numbers
F	The space of weak learners
f_k	k^th weak learner
β	Boosting rate
$(\bar{y_i})_m$	Output of m^th weak learner for i^th datapoint
$\mathcal{E}(x_i)$	Error for i^th datapoint

rate to the feature vector, and then finally perform the classification task using the modified data. For the regression task, we introduce a BNN. Fig. 1 depicts the flowchart of our proposed work for mental stress detection. The successive sub-sections contain a detailed description of the various steps of the proposed method. Tab. 3 tabulates the various symbols that are used in the successive sub-sections.

A. DATA PRE-PROCESSING

WESAD dataset is a huge dataset as each subject's data is recorded at a quite high frequency of 700 Hz. We convert the time series data to a discrete form for which we use a sliding window of 2 seconds. The window is shifted with steps of 0.25 seconds. A lot of samples with redundant values are generated as a result of this discretization step. Hence, for computational feasibility as well as removing the redundant data, the data is aggregated using a window of 5 seconds, and the values in that interval are replaced with their arithmetic mean. This step is unlikely to cause any significant data loss as mental stress levels will not see a drastic change in such a short interval. The next step in our work is to balance the training dataset as the original dataset is unbalanced with the three affective states in an approximate proportion of 1 (amusement): 3 (neutral): 2 (stress). A model trained on such an unbalanced dataset is prone to some biasing. Hence the dataset is balanced by duplicating data corresponding to affective states having a lesser share in the dataset. We divide data points with labels neutral (baseline) and stress into three and two parts respectively so that each part almost consists of the same number of data points as the data points with amusement labels. Then we make three sets by combining one part from each label as depicted in Fig. 2.

Finally, we concatenate the three sets. We balance the dataset in this manner to increase its size as opposed to reducing it using some standard data clipping methods for dataset balancing. After these pre-processing steps, all the labels in the dataset are approximately in the same proportion.

The above steps is followed by dataset normalization to bring the values of biomarkers into the same range. We use min-max scaling for normalization to bring the value of each biomarker between 0 and 1. Mathematically normalisation is expressed as Eq. 1

$$X_{in} = \frac{X_i - X_{minimum}}{X_{maximum} - X_{minimum}} \tag{1}$$

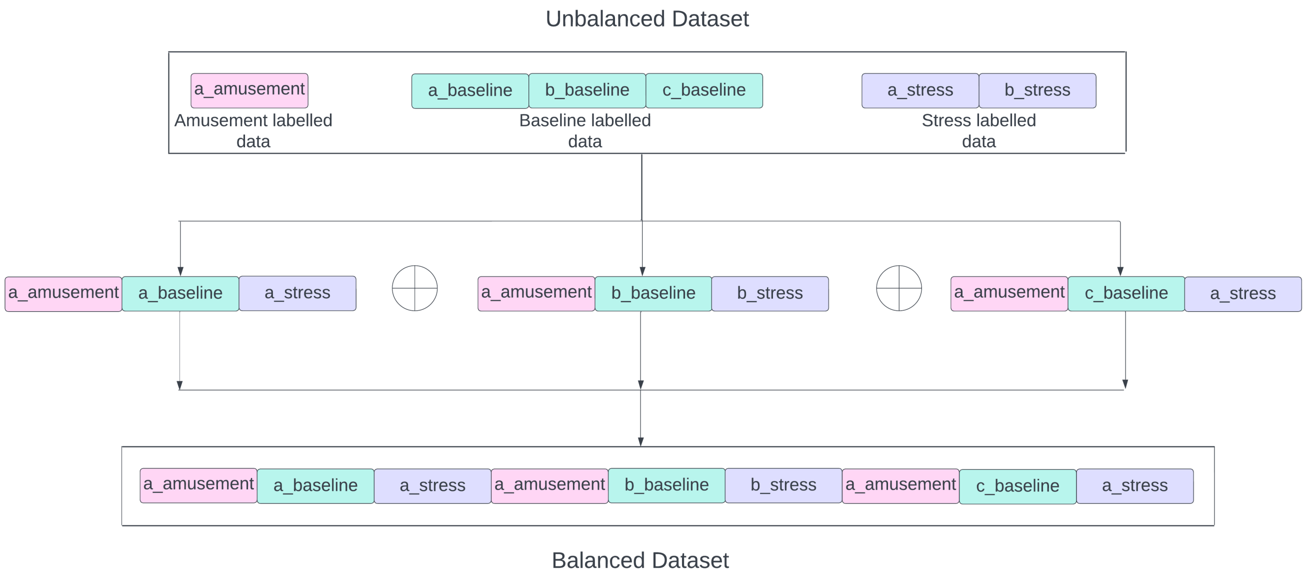


FIGURE 2. Segmenting dataset in multiple parts such that each part contains almost the same number of data points and combining them in a manner that balances the dataset.

where $X_{minimum}$ is the minimum of X, $X_{maximum}$ is the maximum of X and X_{in} is the normalised value of i^{th} data point X_i of X.

The pre-processed dataset after the above steps consists of 7695 samples in the train set and 1342 samples in the test set having an almost equal share of samples with the three affective states in the train set.

B. MODEL TRAINING

The goal of this work is to develop a model for mental stress detection that can be embedded within a smartwatch. Hence we use only the biomarkers that are measured by a smartwatch. Our proposed work consists of two phases. The first one is a regression model that predicts respiration rate using the physiological biomarkers measured by smartwatches and the second one uses a classifier that detects mental stress levels (three affective states viz amusement (assigned class 0), neutral (assigned class 1), stress (assigned class 2)). For the regression task, we introduce a BNN inspired by the works of Friedman [54] and Liu and Badirli [55]. In their proposed boosting approach Liu et al. use an ANN as weak learners. To train a particular weak learner they combine the prediction results from the penultimate layer of the previous ANN (weak learner) with the original input features. The final output is calculated by combining output from each weak learner. In sub-section IV-B1 we explain our approach of creating a BNN. A simple ANN and an autoencoder were also considered for regression tasks but the proposed BNN outperformed both of them. The second sub-model is an ANN that is used as a classifier for detecting mental stress whose architectural details are included in section IV-B2.

1) BOOSTING NEURAL NETWORK: FOR PREDICTING RESPIRATION RATE

Boosting algorithms and neural networks both provide good regression power individually. Traditional boosting techniques use a number of weak learners, typically decision trees, and leverages their combined power for regression and classification tasks. We use shallow neural networks instead of decision trees which paves the way for deep learning's modeling capabilities to be used with boosting techniques. The experimental results, discussed in section V-A, support the above claim by showing better results for BNN as compared to a simple ANN.

We use shallow neural networks having two layers as the weak learners for the boosting model. A series of 6 weak learners (the number of weak learners is a hyperparameter) are used for boosting. In the boosting algorithm, each weak learner predicts the error between the predicted value and the actual value of the target. We initialize the predicted values by assigning them a value equal to the mean of the actual values of respiration rates from the WESAD dataset. We train the model by predicting and eliminating the residual prediction error at each step using weak learners. The rate of learning is controlled by a user-defined hyperparameter, namely the boosting rate. We keep the boosting rate as the reciprocal of the number of weak learners to allow each weak learner to mitigate a fraction of the prediction error such that the overall model has minimal error. We formally present the mathematical formulation of our model below.

Considering a dataset D with n samples having d dimensional feature space

$$D = \{(x_i, y_i)_{i=1}^{n} \mid x_i \in \mathbb{R}^d,\ y_i \in \mathbb{R}\} \tag{2}$$

Then the output of BNN will be given using K additive function given by Eq. 3

$$\bar{y_i} = \sum_{k=0}^{K} \beta_k f_k(x_i), \ f_k \in F \tag{3}$$

where F is the space of weak learners, the neural networks, β_k is the boosting rate K denotes the number of weak learners and each function f_k represents an independent, shallow neural network. The model's prediction is initialized with the mean of actual target values, i.e. $f_0 = \frac{\Sigma_{i=1}^{n}(x_i)}{n}$.

For each weak learner, the target value provided is the difference between the actual value y_i and the intermediate output of BNN given by eq 4. It can be viewed as the overall error value for the BNN model. Mathematically for the m^{th} weak learner the target value provided or alternatively the error function can be represented by eq 5

$$(\bar{y_i})_m = \sum_{k=0}^{m} \beta_k f_k(x_i) \tag{4}$$

$$(\mathcal{E}(x_i))_m = y_i - (\bar{y_i})_m \tag{5}$$

As the training progresses, with the aid of each weak learner the error $\mathcal{E}(x_i)$ is gradually minimized. Fig. 3 gives a diagrammatic representation of the BNN model adopted in this work. In the diagram, D represents the dataset, x represents the feature vector and y represents the target value. The diagram shows the k^{th} and $(k+1)^{th}$ weak learner of the model. The k^{th} layer accepts x as the input and $(y - \bar{y}_{k-1})$ as the target value. $f_k(x)$ denotes the output of the k^{th} weak learner which is then multiplied with boosting rate β_k for the k^{th} weak learner. Finally, $\bar{y}_{k-1}$ is added to that product, and the combined sum forms $\bar{y_k}$ which is passed on to the next weak learner, in the same manner, $\bar{y}_{k-1}$ was passed to k^{th} weak learner.

2) ARTIFICIAL NEURAL NETWORK: FOR CLASSIFICATION

The mental stress detection problem can be viewed as a multi-class or multi-label classification problem. Three different affective states viz. Amusement (class 0), Baseline or neutral (class 1), and Stress (class 2) constitute the target classes. For the classification task, we employ an ANN. For choosing the optimal architecture of the ANN we perform extensive hyperparameter tuning. The proposed ANN architecture is four layers deep with 64, 128, 64, and 1 being the number of neurons in each layer respectively. The ANN uses the ReLU (Rectified Linear Unit) activation function at each layer, Adam optimizer for optimized gradient descent, and a learning rate of 0.1.

C. OVERALL PROPOSED ALGORITHM

Algorithm 1 summarizes the steps of our proposed method. In the algorithm, the variables followed by parentheses denote a function and the ones without them represent data. Physiological biomarkers, namely ECG signals, EDA signals, EMG signals, body temperature, respiration rate, three-axis acceleration, and corresponding labels for affective state, namely amusement, neutral, and stress, form the input of our algorithm. The output of the algorithm consists of evaluated performance metrics. The algorithm starts with pre-processing of the dataset (lines 1 to 4 of algorithm 1). The time series dataset is discretized to generate discrete data samples, aggregated using a suitable window W, balanced to prevent biasing of the model, and finally normalized to improve model performance. Since the WESAD dataset already includes respiration rate, the respiration rate is extracted from the feature set of the pre-processed dataset. Lines 6 and 7 of algorithm 1, depict the regression task using BNN, details of which are presented in Section IV-B1. Boosting Neural Network() takes dataset (*wResp*; stands for dataset without respiration rate) as a parameter. Its other parameters include the number of weak learners(w), boosting rate(β), number of iterations for each weak learner(n), and model to be used as a weak learner (m). The BNN model is depicted as *BNNmodel*() in the algorithm and produces the respiration rate denoted by *resp*. The predicted respiration rate is combined with the feature set of *wResp* and the modified data set (*respD*) is fed as input to the classifier (line 9 of algorithm 1), for which we used a four-layer neural network. Finally, in line 10 of the algorithm, the various performance metrics are evaluated.

Algorithm 1 Algorithm for Our Proposed Resp-BoostNet

Input: ECG signals, EDA signals, EMG signals, body temperature, respiration rate and three-axis acceleration (from WESAD dataset), D; Labels (amusement, neutral, stress), L

Output: Evaluated performance metrics

1. $d \leftarrow$ Discretize the time series dataset D
2. $agg \leftarrow$ Aggregate the data points using a window W
3. $bal \leftarrow$ Bal(agg), to balance the dataset
4. $norm \leftarrow$ Norm(bal), to normalise the dataset
5. $wResp \leftarrow$ Extract respiration rate from feature set of dataset $norm$
6. $BNNmodel() \leftarrow$ Boosting Neural Network($wResp$, w, β, n, m)
7. $resp \leftarrow BNNmodel(wResp)$, To generate respiration rate
8. $respD \leftarrow$ To combine predicted respiration rate $resp$ within feature set of dataset $wResp$
9. $pred \leftarrow$ Classify($respD$), to predict stress level using $respD$
10. $perf \leftarrow$ Eval($pred$, L), to evaluate different performance metrics

return $perf$

D. HYPERPARAMETER TUNING

Hyperparameters are tunable parameters that have a significant impact on the performance of a deep learning based model. Hence selection of suitable hyperparameters becomes an essential task. For this work, we used the Random Search technique to find the most appropriate hyperparameters. We iterated over different set of values

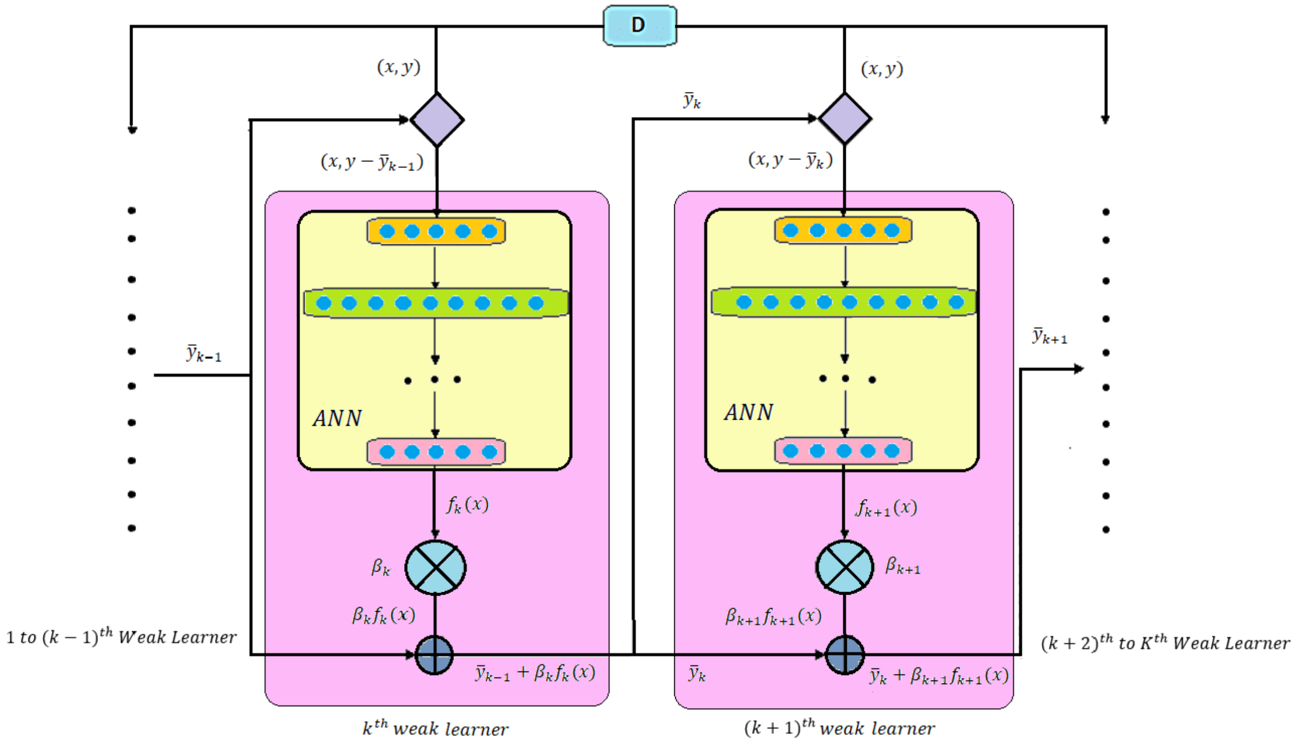


FIGURE 3. The architecture of boosting neural network introduced in this study.

for the hyperparameters and selected the ones that resulted in the highest accuracy of the model. For the regressor, i.e. the BNN, we tried different sets of number of weak learners (shallow neural network), boosting rate, number of iterations for each weak learner and optimizer to be used during gradient descent. For the classifier, i.e. the ANN, we tried different architectures (number of layers and number of neurons in each layer) and different learning rates. Tab. 4 lists the values of different hyperparameters chosen by us after performing extensive hyperparameter tuning for the regressor and classifier.

TABLE 4. The various hyperparameters for the regressor and the classifier of the proposed work.

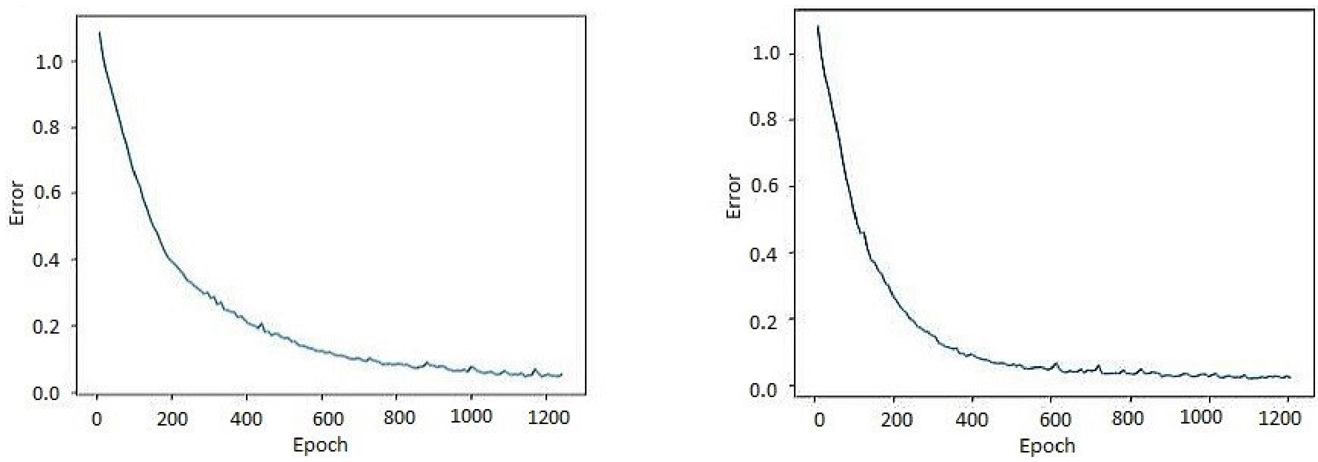
Regressor hyperparameters (BNN)	
Hyperparameter	**Description or value**
Number of weak learners (ANN)	6
Boosting rate	0.166
Iterations for each weak learner	100
Optimizer	Adam
Classifier hyperparameters (ANN)	
Hyperparameter	**Description or value**
Number of layers	4
Neurons in each layer	64, 128, 64, 3
Number of iterations	1250
Learning rate	0.001

V. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we present a comprehensive set of experiments and the performance of the model to demonstrate the effectiveness of our proposed method for mental stress detection. This section also contains a comparative study involving various machine learning approaches namely, Linear and RBF, SVM, DT (Decision Tree), RF (Random Forest), AdaBoost, GNB (Gaussian Naive Bayes), and our proposed ANN as a classifier. We also made a performance comparison of our proposed method with different contemporary techniques that used the WESAD dataset, namely, Sirrtola [12], Kumar et. al. [24], Kumar et. al. [38], Ninh et. al. [56], Iqbal et. al. [51], Indikawati and Winiarti [52] and Gjoreski et. al. [30] for mental stress detection. Section V-B deals with the first part of experimental analysis, i.e. the comparison with different baseline techniques, while section V-C deals with the latter one and presents a detailed performance comparison with various contemporary techniques listed above. We have used Python programming language for the implementation of the code for the proposed method. The work primarily uses Python's library for machine learning and deep learning in PyTorch, and some other libraries such as sklearn, pandas, and numpy. The experimental results are discussed below.

A. PERFORMANCE OF AUXILIARY TASKS

The first series of experiments involved in our work was assessing the importance of each biomarker in detecting mental stress levels. As discussed in section IV, we trained a



(a) (b)

FIGURE 4. Error vs Epoch plot for the proposed model. Figure (a) shows plot for the case when respiration rate is not present in the dataset and figure (b) shows plot for the case when predicted respiration rate is included in the dataset.

TABLE 5. Performance of neural network on WESAD dataset with different sets of biomarkers as feature.

Biomarkers	Accuracy%
Acc, Temp, Resp, EDA, ECG, EMG	95.4
Temp, Resp, EDA, ECG, EMG	89.4
Acc, Resp, EDA, ECG, EMG	91.5
Acc, Temp, Resp, ECG, EMG	92.1
Acc, Temp, Resp, EDA, EMG	92.4
Acc, Temp, Resp, EDA, ECG	90.7
Acc, Temp, EDA, ECG, EMG	91.6

TABLE 6. MSE values for autoencoder, artificial neural network, and boosting neural network when used for regression task using WESAD dataset.

Regression model	MSE
Autoencoder	0.15
ANN	0.038
BNN	0.035

neural network on the WESAD dataset multiple times, each time removing one of the biomarkers from the dataset to observe its impact on the model's performance. The accuracy of the neural network for each case is enlisted in Tab. 5. It is evident from this table that respiration plays a vital role in detecting mental stress.

For predicting respiration rate we proposed a BNN. Tab. 6 shows the MSE values for an autoencoder, an ANN, and the proposed BNN when trained on the WESAD dataset. It is evident from MSE values that the proposed BNN for regression outperforms the autoencoder and a simple ANN.

B. COMPARISON WITH VARIOUS BASELINE TECHNIQUES

This section contains the comparison of various baseline machine learning techniques and our proposed method, Resp-BoostNet on the WESAD dataset. We trained each model thrice to compare their performance in different scenarios which are; when the respiration rate is removed from the feature set of the WESAD dataset when the respiration rate is used from the WESAD dataset itself and finally when the WESAD dataset's respiration rate data is replaced by respiration rate predicted by the BNN based regressor (Tab. 7). Fig. 4 shows the Error versus Epoch plots for the proposed model. The plots indicate that the inclusion of respiration rate improved the model's performance which is evident from the fact that Fig 4(b) has a better learning curve.

To assess the generalizability of the models and more accurate performance estimation, we also performed k-fold cross-validation. The optimum value of k was found to be 8 and all the reported metrics are evaluated for this value of k. The data from Tab. 7 clearly shows that our proposed method has a better performance as compared to the various baseline methods including the traditional boosting method. Also, the accuracy of the model is significantly improved when the respiration rate is incorporated into the feature set of the WESAD dataset. Furthermore, the performance of the classifier when it is trained on a dataset containing originally measured respiration rate and when the dataset has predicted values is almost similar. Tab. 7 shows the performance of the proposed method, along with other baseline techniques, on the entire WESAD dataset, containing data from 15 subjects. The classifier was also trained on individual data of the 15 subjects and Fig. 5 shows the performance of the classifier for each subject when the respiration rate is removed from the WESAD dataset (red plot) and when predicted respiration rate is added to feature set of WESAD dataset (green plot). The figure clearly shows that the inclusion of respiration rate in the feature set improves the classifier's performance for most of the subjects. The classifier used in the proposed work was selected after performing extensive hyperparameter tuning as discussed in IV-D. Fig. 6 shows a comparison of the performance of artificial neural networks with different learning rates on the WESAD dataset without the respiration

TABLE 7. Performance comparison of our proposed Resp-BoostNet framework with different baseline machine learning techniques with and without using respiration rate.

		Methods						
		Linear SVM	RBF SVM	DT	RF	AdaBoost	GNB	**Resp-BoostNet**
Without Respiration Rate	Accuracy (%)	41.4 ± 1.6	70.3 ± 2.0	71.6 ± 2.0	77.3 ± 1.6	76.9 ± 1.0	44.1 ± 0.7	88.5 ± 1.0
	Precision (%)	42.9 ± 1.6	78.4 ± 1.5	79.1 ± 1.6	80.9 ± 1.6	78.3 ± 1.2	49.6 ± 0.8	91.0 ± 0.6
	Recall (%)	41.5 ± 1.5	72.2 ± 1.8	71.5 ± 2.0	77.1 ± 1.6	77.3 ± 1.0	44.9 ± 0.7	88.6 ± 0.9
	F1 Score (%)	36.0 ± 1.5	64.9 ± 2.3	66.7 ± 2.3	74.1 ± 1.8	77.8 ± 1.3	39.3 ± 0.9	90.4 ± 0.7
With WESAD's Respiration Rate	Accuracy (%)	41.4 ± 1.6	70.0 ± 2.1	71.8 ± 2.1	78.6 ± 0.7	77.3 ± 1.2	44.1 ± 0.7	$\mathbf{94.1 \pm 0.6}$
	Precsion (%)	42.9 ± 1.6	76.9 ± 1.8	79.2 ± 1.6	78.1 ± 1.7	80.1 ± 1.7	49.6 ± 0.8	$\mathbf{95.3 \pm 0.5}$
	Recall (%)	41.5 ± 1.5	71.6 ± 1.9	72.4 ± 2.0	73.6 ± 1.7	79.7 ± 1.2	44.9 ± 0.7	$\mathbf{95.3 \pm 0.5}$
	F1 Score (%)	36.0 ± 1.5	63.3 ± 2.4	67.2 ± 2.3	76.8 ± 2.0	79.8 ± 0.5	39.3 ± 0.9	$\mathbf{94.9 \pm 0.6}$
With Predicted Respiration Rate	Accuracy (%)	41.4 ± 1.5	70.0 ± 2.1	69.9 ± 2.2	78.9 ± 0.5	77.3 ± 1.0	45.3 ± 0.7	$\mathbf{94.0 \pm 0.3}$
	Precision (%)	42.9 ± 1.6	76.9 ± 1.8	75.2 ± 2.0	76.1 ± 1.2	78.1 ± 1.2	46.1 ± 1.0	$\mathbf{95.5 \pm 0.7}$
	Recall (%)	41.5 ± 1.5	71.6 ± 1.9	70.7 ± 2.1	81.6 ± 1.7	82.4 ± 0.9	46.2 ± 0.7	$\mathbf{97.0 \pm 0.3}$
	F1 Score (%)	36.0 ± 1.5	66.3 ± 2.4	66.8 ± 2.4	78.8 ± 1.7	80.8 ± 0.7	40.0 ± 0.9	$\mathbf{96.9 \pm 0.6}$

TABLE 8. Performance comparison of the proposed Resp-BoostNet framework with different contemporary techniques for mental stress detection (Model's accuracy when trained on data for all 15 subjects of WESAD dataset).

Methods	**Accuracy%**
Gjoreski et al. [30]	85
Siirtola et al. [12]	87.4
Indikawati et al. [52]	92
Kumar et al. [24]	86.8
Sharma et. al. [38]	87.7
Iqbal et al. [51]	85.71
Tran et. al. [56]	87.5
Proposed Resp-BoostNet	**94.0**

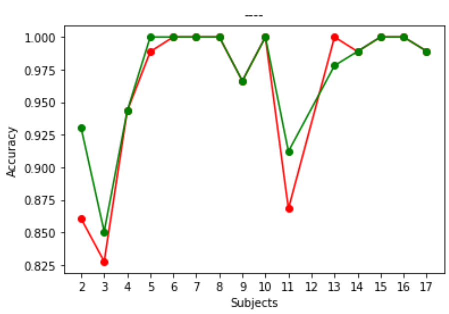


FIGURE 5. Accuracy for each subject for dataset without respiration rate (red plot) and with predicted respiration rate data (green plot).

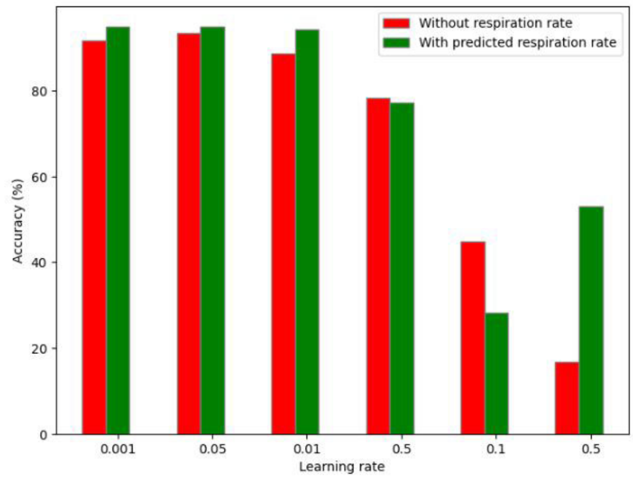


FIGURE 6. A comparison of classification accuracy of artificial neural network for different learning rates with and without predicted respiration rate in the dataset.

rate and with the predicted respiration rate in its feature set. Smaller learning rates make the learning process very time-consuming and larger learning rates affect the model's performance hence, the search range was kept between 0.001 to 0.5. The optimal performance was obtained at a learning rate of 0.001

Hence, Tab. 7 clearly shows that our proposed model outperforms all other baseline machine-learning algorithms. AdaBoost-based classifier is the second best performer after the proposed model followed by the Random Forest classifier while Linear SVM is the worst-performing technique because of its limited ability to understand multi-dimensional data.

C. COMPARISON WITH VARIOUS CONTEMPORARY TECHNIQUES

The previous section contained the comparison between various baseline machine learning techniques and the proposed method Resp-BoostNet. This section compares the performance of our proposed method, Resp-BoostNet with several existing contemporary techniques for mental stress detection. Tab. 8 enlists the average accuracy of different techniques and our method for the 15 subjects in the WESAD dataset. The results listed in Tab. 8 demonstrate that the proposed methodology outperforms other methods. The inclusion of predicted respiration rate in the biomarker set increases the number of biomarkers used which explains the improved performance of our method as compared to some of the works (for instance [52]). But some of the related works use the same set of biomarkers as our work but still have comparatively lower performance. The improved performance in those cases could be attributed to the efficient pre-processing of the dataset performed.

VI. CONCLUSION

Mental stress adversely affects the mental as well as physical health of a person and can lead to serious complications in the long run. This forms the pretext for developing an efficient technique for mental stress detection. However, predicting an individual's mental stress level from high-frequency, multivariate real-time biomarkers using multiple wearable devices is a complex process. The existing works either use sophisticated devices for measuring different physiological biomarkers to detect stress with higher accuracy or detect stress with comparatively lesser accuracy using limited physiological biomarkers measured with help of convenient measurable devices such as smartwatches. We tried to narrow this gap by providing a method that can detect stress with high accuracy using a smartwatch which can be conveniently worn. We observed that the respiration rate is an important biomarker for mental stress detection. The proposed model was able to achieve a high detection accuracy and outperform several contemporary methods using the limited biomarkers combined with respiration rate predicted using those limited biomarkers measured by a smartwatch. The study encountered a limitation due to the scarcity of publicly available datasets containing a comprehensive set of biomarkers. Moreover, the proposed model was trained on a single dataset containing data from 15 subjects and hence the trained model is prone to make generalized predictions. As each person possesses a distinct physiology, research focused on the development of personalized detection models is expected to produce significant improvements in results. A semi-supervised model for fine-tuning the proposed mental stress detection model, trained on data from multiple sources, in order to capture personalized physiological changes in response to stress conditions can be a potential direction for future research.

REFERENCES

[1] M. Chesnut, S. Harati, P. Paredes, Y. Khan, A. Foudeh, J. Kim, Z. Bao, and L. M. Williams, ''Stress markers for mental states and biotypes of depression and anxiety: A scoping review and preliminary illustrative analysis,'' *Chronic Stress*, vol. 5, Jan. 2021, Art. no. 247054702110003, doi: 10.1177/24705470211000338.

[2] S. Segerstrom and G. Miller, ''Psychological stress and the human immune system: A meta-analytic study of 30 years of inquiry,'' *Psychol. Bull.*, vol. 130, no. 4, p. 601, 2004.

[3] A. Angeli, M. Minetto, A. Dovio, and P. Paccotti, ''The overtraining syndrome in athletes: A stress-related disorder,'' *J. Endocrinolog. Invest.*, vol. 27, no. 6, pp. 603–612, Jun. 2004.

[4] A. R. Chauhan and S. Kumar, ''Analysing effectiveness of different physiological biomarkers in detecting stress,'' in *Proc. IEEE World Conf. Appl. Intell. Comput. (AIC)*, vol. 16, Jul. 2023, pp. 71–75.

[5] E. Oetter. (2011). *Heart Rate Variability Research Review*. [Online]. Available: http://www.8weeksout.com/2011/12/05/heart-rate-variability-research-review/

[6] B. Javed, A. Sarwer, E. B. Soto, and Z.-U.-R. Mashwani, ''The coronavirus (covid-19) pandemic's impact on mental health,'' *The Int. J. Health Planning Manage*, vol. 35, no. 5, pp. 993–996, Sep. 2020.

[7] S. Cohen, T. Kamarck, and R. Mermelstein, ''A global measure of perceived stress,'' *J. Health Social Behav.*, vol. 24, no. 4, p. 385, Dec. 1983.

[8] N. Parameswaran and D. Venkataraman, ''A computer vision based image processing system for depression detection among students for counseling,'' *Indonesian J. Electr. Eng. Comput. Sci.*, vol. 14, no. 1, p. 503, Apr. 2019.

[9] J. A. Healey and R. W. Picard, ''Detecting stress during real-world driving tasks using physiological sensors,'' *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 2, pp. 156–166, Jun. 2005.

[10] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, ''DEAP: A database for emotion analysis; using physiological signals,'' *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18–31, Jan. 2012.

[11] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, ''Introducing WESAD, a multimodal dataset for wearable stress and affect detection,'' in *Proc. 20th ACM Int. Conf. Multimodal Interact.* New York, NY, USA: ACM, Oct. 2018, pp. 400–408, doi: 10.1145/3242969.3242985.

[12] P. Siirtola, ''Continuous stress detection using the sensors of commercial smartwatch,'' in *Proc. Adjunct ACM Int. Joint Conf. Pervasive Ubiquitous Comput. Proc. ACM Int. Symp. Wearable Comput.* New York, NY, USA: ACM, Sep. 2019, pp. 1198–1201, doi: 10.1145/3341162.3344831.

[13] M. Gjoreski, H. Gjoreski, M. Luštrek, and M. Gams, ''Continuous stress detection using a wrist device: In laboratory and real life,'' in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput., Adjunct*. New York, NY, USA: ACM, Sep. 2016, pp. 1185–1193, doi: 10.1145/2968219.2968306.

[14] A. Salazar-Ramirez, E. Irigoyen, R. Martinez, and U. Zalabarria, ''An enhanced fuzzy algorithm based on advanced signal processing for identification of stress,'' *Neurocomputing*, vol. 271, pp. 48–57, Jan. 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S092523121731216X

[15] A. Bannore, T. Gore, A. Raut, and K. Talele, ''Mental stress detection using machine learning algorithm,'' in *Proc. Int. Conf. Electr., Comput., Commun. Mechatronics Eng. (ICECCME)*, Oct. 2021, pp. 1–4.

[16] P. Schmidt, A. Reiss, R. Duerichen, and K. Van Laerhoven, ''Wearable affect and stress recognition: A review,'' 2018, *arXiv:1811.08854*.

[17] D. Widjaja, M. Orini, E. Vlemincx, and S. Van Huffel, ''Cardiorespiratory dynamic response to mental stress: A multivariate time-frequency analysis,'' *Comput. Math. Methods Med.*, vol. 2013, May 2013, Art. no. 451857. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1155/2013/451857

[18] I. Homma and Y. Masaoka, ''Breathing rhythms and emotions,'' *Experim. Physiol.*, vol. 93, no. 9, pp. 1011–1021, Sep. 2008.

[19] R. Jerath, M. W. Crawford, V. A. Barnes, and K. Harden, ''Self-regulation of breathing as a primary treatment for anxiety,'' *Appl. Psychophysiol. Biofeedback*, vol. 40, no. 2, pp. 107–115, Jun. 2015, doi: 10.1007/s10484-015-9279-8.

[20] H. Qin, X. Ma, Y. Ding, X. Li, Y. Zhang, Z. Ma, J. Wang, J. Luo, and X. Liu, ''BiFSMNv2: Pushing binary neural networks for keyword spotting to real-network performance,'' *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 8, pp. 10674–10686, Aug. 2024.

[21] H. Qin, Y. Ding, X. Zhang, J. Wang, X. Liu, and J. Lu, ''Diverse sample generation: Pushing the limit of generative data-free quantization,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 10, pp. 11689–11706, Oct. 2023.

[22] H. Qin, X. Zhang, R. Gong, Y. Ding, Y. Xu, and X. Liu, ''Distribution-sensitive information retention for accurate binary neural network,'' *Int. J. Comput. Vis.*, vol. 131, no. 1, pp. 26–47, Oct. 2022, doi: 10.1007/s11263-022-01687-5.

[23] H. Qin, M. Zhang, Y. Ding, A. Li, Z. Cai, Z. Liu, F. Yu, and X. Liu, ''BiBench: Benchmarking and analyzing network binarization,'' 2023, *arXiv:2301.11233*.

[24] A. Kumar, K. Sharma, and A. Sharma, ''Genetically optimized fuzzy C-means data clustering of IoMT-based biomarkers for fast affective state recognition in intelligent edge analytics,'' *Appl. Soft Comput.*, vol. 109, Sep. 2021, Art. no. 107525. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1568494621004488

[25] D. Gupta, M. P. S. Bhatia, and A. Kumar, ''Resolving data overload and latency issues in multivariate time-series IoMT data for mental health monitoring,'' *IEEE Sensors J.*, vol. 21, no. 22, pp. 25421–25428, Nov. 2021.

[26] K. Hovsepian, M. al'Absi, E. Ertin, T. Kamarck, M. Nakajima, and S. Kumar, ''CStress: Towards a gold standard for continuous stress assessment in the mobile environment,'' in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.* New York, NY, USA: ACM, Sep. 2015, pp. 493–504, doi: 10.1145/2750858.2807526.

[27] L. D. Sharma, V. K. Bohat, M. Habib, A. M. Al-Zoubi, H. Faris, and I. Aljarah, "Evolutionary inspired approach for mental stress detection using EEG signal," *Expert Syst. Appl.*, vol. 197, Jul. 2022, Art. no. 116634. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417422001233

[28] S. Samarpita and R. Satpathy, "Eeg-based stress detection using k-means clustering method," in *Intelligent Systems and Machine Learning*. Cham, Switzerland: Springer, 2023, pp. 35–43.

[29] W. Sanchez, A. Martínez-Rebollar, and M. Gonzalez-Mendoza, "Towards job stress recognition based on behavior and physiological features," in *Proc. Int. Conf. Ubiquitous Comput. Ambient Intell.*, 2017, pp. 311–322.

[30] M. Gjoreski, M. Luśtrek, M. Gams, and H. Gjoreski, "Monitoring stress with a wrist device using context," *J. Biomed. Informat.*, vol. 73, pp. 159–170, Sep. 2017.

[31] Z. Guo, Y.-J. Wang, B.-S. He, and J. Zhou, "Linc00312 single nucleotide polymorphism as biomarker for chemoradiotherapy induced hematotoxicity in nasopharyngeal carcinoma patients," *Disease Markers*, vol. 2022, pp. 1–9, Aug. 2022.

[32] H. Wang, T. Yang, J. Wu, D. Chen, and W. Wang, "Unveiling the mystery of SUMO-activating enzyme subunit 1: A groundbreaking biomarker in the early detection and advancement of hepatocellular carcinoma," *Transplantation Proc.*, vol. 55, no. 4, pp. 945–951, May 2023.

[33] M. Ciman and K. Wac, "Individuals' stress assessment using human-smartphone interaction analysis," *IEEE Trans. Affect. Comput.*, vol. 9, no. 1, pp. 51–65, Jan. 2018.

[34] K. Yang, C. Wang, Y. Gu, Z. Sarsenbayeva, B. Tag, T. Dingler, G. Wadley, and J. Goncalves, "Behavioral and physiological signals-based deep multimodal approach for mobile emotion recognition," *IEEE Trans. Affect. Comput.*, vol. 14, no. 2, pp. 1082–1097, Jun. 2021.

[35] E. A. Sağbaş, S. Korukoglu, and S. Balli, "Stress detection via keyboard typing behaviors by using smartphone sensors and machine learning techniques," *J. Med. Syst.*, vol. 44, no. 4, p. 68, Feb. 2020. [Online]. Available: https://doi.org/10.1007/s10916-020-1530-z

[36] P. Garg, J. Santhosh, A. Dengel, and S. Ishimaru, "Stress detection by machine learning and wearable sensors," in *Proc. 26th Int. Conf. Intell. User Interfaces*. New York, NY, USA: ACM, Apr. 2021, pp. 43–45, doi: 10.1145/3397482.3450732.

[37] L. Zhu, P. Spachos, P. C. Ng, Y. Yu, Y. Wang, K. Plataniotis, and D. Hatzinakos, "Stress detection through wrist-based electrodermal activity monitoring and machine learning," *IEEE J. Biomed. Health Informat.*, vol. 27, no. 5, pp. 2155–2165, May 2023.

[38] A. Kumar, K. Sharma, and A. Sharma, "Hierarchical deep neural network for mental stress state detection using IoT based biomarkers," *Pattern Recognit. Lett.*, vol. 145, pp. 81–87, May 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167865521000490

[39] H. Yu, T. Vaessen, I. Myin-Germeys, and A. Sano, "Modality fusion network and personalized attention in momentary stress detection in the wild," in *Proc. 9th Int. Conf. Affect. Comput. Intell. Interact. (ACII)*, Sep. 2021, pp. 1–8.

[40] S. Ahmed, M. A. Yousuf, M. M. Monowar, A. Hamid, and M. O. Alassafi, "Taking all the factors we need: A multimodal depression classification with uncertainty approximation," *IEEE Access*, vol. 11, pp. 99847–99861, 2023.

[41] N. Khan and N. Sarkar, "Semi-supervised generative adversarial network for stress detection using partially labeled physiological data," 2022, *arXiv:2206.14976*.

[42] I. Zyma, S. Tukaev, I. Seleznov, K. Kiyono, A. Popov, M. Chernykh, and O. Shpenkov, "Electroencephalograms during mental arithmetic task performance," *Data*, vol. 4, no. 1, p. 14, Jan. 2019. [Online]. Available: http://www.mdpi.com/2306-5729/4/1/14

[43] V. Markova, T. Ganchev, and K. Kalinkov, "CLAS: A database for cognitive load, affect and stress recognition," in *Proc. Int. Conf. Biomed. Innov. Appl. (BIA)*, Nov. 2019, pp. 1–4.

[44] J. Birjandtalab, D. Cogan, M. B. Pouyan, and M. Nourani, "A non-EEG biosignals dataset for assessment and visualization of neurological status," in *Proc. IEEE Int. Workshop Signal Process. Syst. (SiPS)*, Oct. 2016, pp. 110–114.

[45] M. Yadav, M. N. Sakib, E. H. Nirjhar, K. Feng, A. H. Behzadan, and T. Chaspari, "Exploring individual differences of public speaking anxiety in real-life and virtual presentations," *IEEE Trans. Affect. Comput.*, vol. 13, no. 3, pp. 1168–1182, Jul. 2022.

[46] J. Yoon, C. Kang, S. Kim, and J. Han, "D-Vlog: Multimodal vlog dataset for depression detection," in *Proc. AAAI Conf. Artif. Intell.*, 2022, vol. 36, no. 11, pp. 12226–12234. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/21483

[47] J. Gratch, R. Arstein, G. Lucas, G. Stratou, S. Scherer, A. Nazarian, R. Wood, J. Boberg, D. DeVault, S. Marsella, D. Traum, A. Rizzo, and L. Morency, "The distress analysis interview corpus of human and computer interviews," in *Proc. LREC*, 2014, pp. 3123–3128.

[48] F. Ringeval, B. Schuller, M. Valstar, N. Cummins, R. Cowie, L. Tavabi, M. Schmitt, S. Alisamir, S. Amiriparian, E.-M. Messner, S. Song, S. Liu, Z. Zhao, A. Mallol-Ragolta, Z. Ren, M. Soleymani, and M. Pantic, "AVEC 2019 workshop and challenge: State-of-mind, detecting depression with AI, and cross-cultural affect recognition," in *Proc. 9th Int. Audio/Visual Emotion Challenge Workshop*, Oct. 2019, pp. 3–12.

[49] H. Cai, Z. Yuan, Y. Gao, S. Sun, N. Li, F. Tian, H. Xiao, J. Li, Z. Yang, X. Li, Q. Zhao, Z. Liu, Z. Yao, M. Yang, H. Peng, J. Zhu, X. Zhang, G. Gao, F. Zheng, R. Li, Z. Guo, R. Ma, J. Yang, L. Zhang, X. Hu, Y. Li, and B. Hu, "A multi-modal open dataset for mental-disorder analysis," *Sci. Data*, vol. 9, no. 1, p. 178, Apr. 2022, doi: 10.1038/s41597-022-01211-x.

[50] L. Mou, C. Zhou, P. Zhao, B. Nakisa, M. N. Rastgoo, R. Jain, and W. Gao, "Driver stress detection via multimodal fusion using attention-based CNN-LSTM," *Expert Syst. Appl.*, vol. 173, Jul. 2021, Art. no. 114693. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417421001342

[51] T. Iqbal, P. Redon-Lurbe, A. J. Simpkin, A. Elahi, S. Ganly, W. Wijns, and A. Shahzad, "A sensitivity analysis of biophysiological responses of stress for wearable sensors in connected health," *IEEE Access*, vol. 9, pp. 93567–93579, 2021.

[52] F. I. Indikawati and S. Winiarti, "Stress detection from multimodal wearable sensor data," *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 771, no. 1, Mar. 2020, Art. no. 012028, doi: 10.1088/1757-899x/771/1/012028.

[53] V. Randazzo, J. Ferretti, and E. Pasero, "ECG WATCH: A real time wireless wearable ECG," in *Proc. IEEE Int. Symp. Med. Meas. Appl. (MeMeA)*, Jun. 2019, pp. 1–6.

[54] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001.

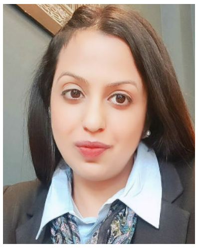
[55] S. Badirli, X. Liu, Z. Xing, A. Bhowmik, K. Doan, and S. S. Keerthi, "Gradient boosting neural networks: GrowNet," 2020, *arXiv:2002.07971*.

[56] V.-T. Ninh, S. Smyth, M.-T. Tran, and C. Gurrin, "Analysing the performance of stress detection models on consumer-grade wearable devices," 2022, *arXiv:2203.09669*.

SANJAY KUMAR received the Ph.D. and M.Tech. degrees in computer applications from Indian Institute of Technology (IIT), New Delhi, India. Previously, he with the National Informatics Centre, Government of India as Scientist-B. He is currently an Assistant Professor with the Department of Computer Science and Engineering, Delhi Technological University (DTU), New Delhi. He has published numerous articles and proceeding papers in reputed journals and conferences. Notably, more than 25 of his articles have been published in top-rated SCI/SCIE journals, reflecting the quality and impact of his work. His research interests include machine learning, AI, social network analysis, NLP, and the design and analysis of algorithms. He was a recipient of the "Commendable Research Award for Excellence in Research" awarded by DTU for the last three consecutive years, namely, in 2020, 2021, and 2022. His name has been included in the "Top 2% scientist of the world" list in the category of Image Processing and AI by Stanford University, USA, in 2023. In addition to his research contributions, he actively participates in the academic community as a reviewer for various journals associated with renowned publishers like Springer and Elsevier

ANSHUMAN RAJ CHAUHAN received the Bachelor of Technology degree in computer science and engineering from Delhi Technological University, in 2023, where he demonstrated a keen interest in the realm of deep learning, health informatics, and NLP. He currently holds the position of a Software Engineer with Google, where he contributes to the cutting-edge Cloud Spanner database within Google Cloud.

AKHIL received the Bachelor of Technology degree from Delhi Technological University, in 2023, where he demonstrated a keen interest in the realm of deep learning. Currently, he is a Software Engineer with Chegg, a leading education technology company, where he has been actively contributing to projects involving large language models. His main research interest includes deep learning. His awards and honors include the Excellence in Education Awards, in 2018.



AKSHI KUMAR (Senior Member, IEEE) is currently a Senior Lecturer and the Director-Post Graduate Research (PGR) of the Department of Computing at the Goldsmiths, University of London, U.K. She was endorsed by the Royal Academy of Engineering, U.K., as an Exceptional promise in the field of artificial intelligence/data science, in 2022. She has received nine research awards for Excellence in Research from various national and international organizations. Her name has been included in the "Top 2% scientist of the world" list by Stanford University, USA, in 2023, 2022, and 2021. She has published more than 95 peer-reviewed journal articles, more than 65 conference papers with five best paper awards, and two patents with the Indian Patent Office. She has been serving as an associate editor and guest editor in various high-impact journals with reputed publishers. Her research interests include affective computing, social network and media analytics, NLP, and AI for pervasive healthcare.

GUANG YANG (Senior Member, IEEE) is currently an Associate Professor (Senior Lecturer) with the Bioengineering Department and Imperial-X, Imperial College London. He holds a UKRI Future Leaders Fellowship and serves as an Honorary Senior Lecturer with the School of Biomedical Engineering and Imaging Sciences, King's College London. His research group is dedicated to developing novel and translational techniques for imaging and biomedical data analysis. The group's focus encompasses research and development in data-driven fast imaging, data harmonization, data synthesis, federated learning, explainable AI, and AI in drug discovery. Currently, his work spans a wide range of clinical applications in cardiovascular disease, lung disease, and oncology. For more information about his Laboratory, visit: https://www.yanglab.fyi/.