

Machine Learning for Stress Prediction: A LightGBM-Based Approach Using Bayesian Optimization

Mohammad Rabib Uddin*, Md. Saniat Rahman Zishan, Shameem Ahmad and Upama Dev

Abstract—Stress is an important physiological and psychological reaction which, when prolonged, has a negative impact on the health and well-being of humans. Traditional stress assessment methods tend to be subjective and are not scalable in real-time. In this research, a lightweight machine learning framework to predict stress from easily measurable environmental and activity-related features, temperature, humidity, and steps, is introduced. The proposed method is different from the approach that depends on complex physiological signals and uses minimal non-invasive inputs that can be applied in the wearable and IoT field. The stress level is considered as the target variable, and all the features used as input variables are independent from each other to prevent data leakage. There are several machine learning and deep learning models and their performances are measured, such as LightGBM, Support vector machine (SVM), AdaBoost, Gated Recurrent Unit (GRU). Bayesian hyperparameter optimization is used to improve the performance of the model and guarantee strong generalization. Experimental results show that LightGBM outperforms the other models, giving an accuracy of 97% and a balanced precision, recall and F1 score of 95%. The results demonstrate promising predictive performance on the evaluated dataset. However, since the model was validated on a relatively small publicly available dataset, further evaluation using larger and more diverse real-world datasets is required before practical deployment.

Index Terms— Machine Learning, Stress Detection, LightGBM, Bayesian Optimization, Classification Model.

I. INTRODUCTION

STRESS is an ordinary as well as adaptive reaction of human beings to unpleasant or dangerous conditions, also known as flight or flight response. This is an instinctive biological response that helps people deal with immediate challenges, threats or stressors. Stress is good in small doses, as it helps

people attend to tasks, be alert and push through challenges. However, prolonged stress that persists over a long period of time can have a significant negative impact on physical and mental health. Cardiovascular disease (high blood pressure, heart disease), weakened immune system, gastrointestinal disease (GI), anxiety disorders and depression are among the medical problems that have been correlated with chronic stress. These conditions impact on an individual's life, their community life and also have a significant burden on a country's health care systems as well as society. It is only recently that following observations of the effects of unmanaged stress several years after exposure have occurred, making it crucial to recognize stress early to take effective countermeasures [1].

In today's world stress, which has become one of the leading causes of diseases and is prevalent in our fast paced and more demanding society. With more pressure than ever to work more to secure employment in the job market and to pay our bills, it's not surprising we are experiencing high levels of stress more than ever. Technology has evolved at a high speed, reducing our energy and setting new standards for connectivity and this has caused us to lose the balance of work and life. Stress is therefore a silent, silent disease that has become common amongst individuals, organizations and communities. This is because, these days the demand for efficient and accurate Stress Detection systems has been increased by a lot. Early and accurate detection of stress can help to provide timely interventions to support adequate stress management before it develops into chronic stress [2]. This is especially so in settings that are likely to be stressful, such as in the workplace, education institutions and healthcare settings, where stress can impact a person's productivity and decision making both individually and in group settings, which can impact wellbeing. Stress assessment has been achieved by traditional means like questionnaires, interviews and observation studies for many years. These are useful tools to measure self-reported stress, but can be subjective and prone to bias. Self-reports are based on the memory of the person and his/her capacity to talk about everything that happened, and that may be influenced by factors like 'mood memory' and 'social desirability'. Stigma may lead to an underreporting of stress, and an over reporting of stress for attention. Studies used in an observational manner are more variable and prone to error, and rely on the interpretation of the

Mohammad Rabib Uddin serves as Lecturer in the department of Electrical and Electronic Engineering at American International University-Bangladesh. Email: rabib@aiub.edu (Corresponding Author)

Md. Saniat Rahman Zishan serves as a Professor in the department of Electrical and Electronic Engineering at American International University-Bangladesh. Email: saniat@aiub.edu

Shameem Ahmad serves as an Associate Professor in the department of EEE, BSRM School of Engineering, BRAC University.

Upama Dev completed her M.Sc. in Electrical and Electronic Engineering at American International University-Bangladesh. Email: 23-93050-2@student.aiub.edu

researchers or clinicians. This needs a paradigm shift in order to find solutions that can objectively, reproducibly and scalable detect the stress [3].

Because of the lack of objectivity in these traditional methods, and the possibility of bias among the experts, other methods are being explored that will likely lead to more reliable and generalizable measures of stress. This system implementation has been enhanced with the aid of physiological and behavioral indices, which can be utilized to transform the system into a continuous and automatic unobtrusive way to measure stress electronically. These systems can make more objective and non-invasive assessments of stress via the characterization of physiological signals (e.g. cardiovascular activity such as blood pressure, respiratory pattern such as breathing rate and real time kinematics such as facial expression) Blood pressure and the respiratory rate are well-known parameters for the changes during the stress response as the autonomic nervous system counters with stressors by modulating them [4]. Similarly, the facial expressions and micro expressions might be an indicator of a fluctuant mood and hence to explore the psyche of a human being. This was pivotal to the idea of getting physiological data in the real world using wearable technology such as smartwatches and fitness trackers. They also have sensors that can sense numerous physiologic signals that generate a wealth of data, suitable for a variety of stress detection analyses. The automated stress detection systems can detect stress patterns from multiple data sources, thanks to the latest advances in machine learning (ML) and deep learning (DL). They have been shown to be effective in analyzing both the environmental, physiological and behavioral information to improve the accuracy of predictions, and also to monitor real time stress with no invasiveness in healthcare and wearable applications [5]. These technologies can produce advanced algorithms that can detect patterns or relationships in the data these technologies collect, which can be converted into action and utilized to help inform an individual's stress level. It's the automatic systems to detect stress that really change lives and are applied to thousands of applications. These systems can be deployed at work to recognize those employees that might be heading for burnout and, in turn, to enable targeted measures like stress management training, flexible working or mental health services. However, if it finds that the employee is under a lot of stress it should give him a break from his boss or a 'help' in the shape of a counsel service [6].

Usage in Healthcare for example, stress detection can assist health care workers to check people with mental illness that have a high degree of resemblance and consequently observing how well they are progressing throughout treatment. At an individual level, such a system can be used to provide them with individualized feedback and suggestions for further action, allowing them to take action and reduce stress. A device, for example, that is able to detect a more serious situation of physical strain with a person could recommend that they take a break or practice being mindful. These foster coping skills and reduce issues stemming from ongoing stress such as becoming more adapted or a non-negative fixation [7].

It is not just that the multimodal data has not been referenced at any point to improve the accuracy, but also the data from the physiology (e.g. heart rate variability), behavior (e.g. movement acceleration) and the environment (e.g. humidity) has not been referenced at any point to improve the resiliency for detecting stress. With that in mind and even more importantly it appears that all of those little details about where the subject is (home vs workplace) and maybe when (what happens after 5:00 p.m.) it is taken gets into the head of blood pressure. Meanwhile, the voice at least the language-voice appears to offer a better set of metrics for various aspects of life (more settings than moments). This way, it is easier to feel stress and more focused (but perhaps less individualized but flexible at best) interventions can be applied. When the system detects that the user is in an advanced working situation, for example, at an early working meeting, it could offer a breathing exercise or just remind users to take short breaks [8].

If the system can have multiple stress (towards user, noise, poor air quality etc) then it's more robust since it can provide some kind of visualization of the situation and of the individual, so the individual feels understood by the system; Sense the reasoning as to why one can be stressed up, literally from within? Well, that's the missing puzzle piece in the stress detection puzzle, anyway. Once those signals can be explained, then it's not all that difficult to cool it down or deal with it. Hence the present requirement and use of stress detection systems. Older approaches are limited in that respect; tech advances don't have that limitation, and they open up new paths and enable novel approaches to noninvasive real time stress monitoring via automation, here. They might be a big stress management game changer everywhere: at work, in health care and in life. This is the only way people/businesses can minimize stressors and keep them from becoming chronic stressors. Thanks to the evolution of technologies, however, there is hope for a world without the negative impacts of stress on patients if they are diagnosed early and monitored ongoing [9].

By developing more accurate, reliable & personalized testing devices, integration of superior ML/DL techniques and knowledge base from physiological, behavioral and environmental data can help enhancement in accuracy of stress anomaly detection systems. As these technologies advance, they will be more focused on assisting humans to live a more balanced lifestyle in terms of mind and body wellness. Hyperparameter tuning is also a significant aspect when creating a strong stress detection system, optimizing parameters of ML and DL models to utilize them for maximum stress prediction accuracy. Hyperparameters are parameters such as learning rate (stepsize), number of layers, number of units in each layer, regularization parameters to prevent overfitting and kernel functions that control the learning of a model. There are many hyper-parameters for DL models, like CNN for vision, or sequential data for RNN, and in addition to determine their values, you also need to make sure that you don't overfit each hyper-parameter on your training data: for each hyper-parameter, you need to use something like validation set with Adversarial examples before you can generalise on your test set. In case of checking different ensemble methods like

Gradient Boosting Machines (GBMs), Random Forests etc most of the hyperparameters are very essential parameters such as tree depth, number of estimators and learning rate etc. This is the best configuration obtained by exploring a systematic way in the hyperparameter space using Grid Search, Random search and Bayesian Optimization techniques. From the application point of view, the improvement in the evaluation of the hyperparameter tuning can contribute to the enhancement of detection accuracy, and the establishment of robust stress detector systems [4]. With optimized parameters; researchers and practitioners can obtain more resilient models that might be used to face the complexity of the stress-based data, which in turn, will help them develop more effective stress mitigation approaches [10].

The study is not really in-depth, on the physiological aspects like EEG, HRV, GSR etc. as most of the studies are based on. However, it is more about the environmental and activity's weak reminders, or cues, which is somewhat different. The entire purpose is to see if it's possible to detect stress patterns with a few, easily-collected data points, such as those that can be tracked by a wearable and IoT based system, that are suitable for computationally efficient machine learning models.

A. Related Work

The science of stress identification has continued to grow in the last ten years or so, and there are now much more advanced stress detection using showier biosensors, wearable devices and computation techniques. The use of physiological signals and machine learning (ML) and deep learning (DL) models has been suggested in previous research to achieve accurate stress detection results is very effective. Singh et al. proved the feasibility of stress detection through biosensor and also showed how the physiological responses like galvanic skin response (GSR) and heart rate variability (HRV) can be captured [11]. Fernandes et al. (2022) also investigated the effectiveness of the use of blood pressure and GSR as indicators of stress, and they revealed a good association between both indicators. The stress detection systems based on sensors [12]. Alberdi et al. (2020) previously pointed out that multimodal data for stress detection should be employed, and in their review they mentioned a previous stress recognition system in the office environment [13]. They stated that, when several physiological signals are combined, such as HRV, skin temperature, and respiratory patterns, the overall performance of the system tends to become better, but in a more stable manner. In addition, Pluntke et al. (2020) noted the role of HRV in predicting physical and psychological stress in firefighters, based on what is achievable with the use of ML approaches. [14] The same HRV was also touted as a parameter with high practical relevance, as it becomes more relevant in various settings. The overall message in this piece was that although the data is complex and physiological, ML models can handle it, but do not break when the data gets messy. The use of Wearable Technology for Advanced Stress Detection. In addition, Sano and Picard (2021) demonstrated that the use of wearables to collect human motion and physiological data is useful for monitors. And even mobile phones, which are sometimes used to sense stress in the real world, and results so far, look promising [15].

Likewise, Sandulescu et al. (2019) examined wearable physiological sensors and found that they are a relatively useful device for continuous monitoring of stress related biomarkers such as heart rate and electrodermal activity (skin conductance) [16]. A few years later, delmastro et al. (2023) [17] took this idea forward, also in the same direction, of using cognitive training coupled with stress detection for elderly people, and measured it using wearable sensors. Several different methods have been suggested for employing ML and DL advances for stress detection. For instance, Elzeiny and Qaraqe (2022) presented a fairly detailed review of ML techniques, and highlighted their advantages, especially SVMs, Random Forests, and ANN [18]. They also emphasized the need for advanced DL models, such as CNN and RNNs, to handle complex time-series physiological data. Fajrin et al. (2024) developed a fuzzy logic-based stress detection system using physiological parameters, including blood pressure and body temperature, to classify stress levels. Periyaswamy and Balasubramanian [19] suggested the application of fuzzy logic in the detection of stress using various physiological signals, in the analysis of stress caused by battle dress uniform. They demonstrated that using some combination of the physiological parameters with fuzzy logic does lead to a more reliable and accurate classification of stress level [20]. This sort of result basically underlines the feasibility of using signals from brain activity as an additional modality, to help identify stress. However, Zhang et al. (2) explored the concept of stress detection using facial expression analysis, while Jha et al. (2020) discussed an application which is Computer Vision. They extracted facial muscle movements using CNNs and micro-expressions, and fed them into the CNN to perform real-time stress level recognition [21].

Stress measurement is non-invasive, contactless and very valuable in the working environment and in the educational sector. Gjoreski et al. (2021) have designed a stress detection system that takes into account the context, and uses sensors on a smartphone and wearable devices. Information from context like location, activity and time of day were incorporated in the design and this led to higher accuracy in detecting stress in real world scenarios [22]. As well, Wang et al. (2023) studied the relationship between environmental factors and stress states, and even included physiological markers with the streams which were collected using the Internet of Things (IoT) sensors (such as temperature, noise and air quality sensors). They realized that there are factors in the environment that can affect stress levels and that's why the significance of multimodal data integration is relevant [26]. The stress detection system proposed by Kumar et al., (2022) based on Thermal imaging recommends that there will be change in face temperature. After that, they employed DL models to decipher the thermal image pattern, and reported a high accuracy for the stress identification. It is a noninvasive, and private method, to the earlier sensor based methods [27]. Tao et al. (2023) delved deeper into the subject and investigated the application of reinforcement learning (RL) in the realm of adaptive stress detection systems. They did train RL agents to adjust model parameters of the system, but with the use of real-time feedback. Well, at least in the 1st, it worked better and was more flexible in other environment as per the result of their process [28]. The other study on stress detection was done by

Hernandez et al. (2022) which was on social media activity. Natural language processing (NLP) of text and sentiment analysis. Based on the results, it can be concluded that social media data can be used to obtain insightful data on stress level especially to the young people [29].

Nguyen et al. (2023) then presented a stress detection framework in which they performed a fusion of multi-modal data. These were physiological, behavioral, and environmental. They used DL models to combine those different data types and analysed them together and claimed state-of-the-art performance for the stress classification [30]. Gupta et al. (2021) looked at gait analysis for stress detection, and used ML methods to study the gait of people as well as human posture. The results they achieved showed that gait based stress detection is a potential low distraction method of unobtrusive monitoring [31]. Since the data set is small, to achieve better performance of the system, the authors of the above paper, Lee et al. (2023), stepped in the field of transfer learning (TL), which is the use of already trained DL models. Transfer learning appears to have been helpful for them to reduce their reliance on large, annotated databases to some degree, and thus, stress detection is becoming possible [32].

B. Contributions of the Paper

The literature review demonstrates that physiological signals can effectively be combined with more sophisticated classification methods to aid in stress detection – but there are significant gaps that make it hard to get this to become a common practice in the field. Many of the datasets are rather small, and they tend to be very specific in the type of population that they represent, so the models don't generalize well. Not to mention the amount of works that are bound to offline stress detection, which don't really offer on-the-fly processing capabilities, which is very important to the actual use. Also, the method of combining multi modal data, such as physiological and behavioral and contextual information, is still rather narrow and weak in general, which is why the detection systems are not that strong. If the idea is to advance the stress detection technology and make it more realistic in the real world and less laboratory-based, it is important to correct these types of problems.

This paper introduces a ML model that can detect the stress level in a constant and true way with the complementary parameter. The aim of this study is to develop an advanced stress assessment model based on ML that has high accuracy for identification and to solve the drawback of the human stress identification process. It was rigorously tested against accuracy, precision and F1-Score as performance metrics to provide justification of the robustness of the proposed framework. In addition, it is competitive compared to the SOTA methods in its effectiveness and viability for practical stress tracking and intervention paradigms in the real world. The main outcomes of this research are:

- ✓ To develop a machine learning model that can better detect stress using environmental cues and features that evolve with time, which are activity-related
- ✓ To create a stress detection model using different ML and DL models, optimize the performance of the model using a Bayesian hyper parameter tuning, yes, that kind of strategy.

- ✓ To check the proposed framework using performance metrics such as accuracy precision and F1 score.
- ✓ To compare and contrast the model with the best-known models available to demonstrate its superiority.

In the remaining paper, methodology is introduced in Section II with details on the stress detection framework such as data description and preparation, and the architecture of the ML and DL models proposed. Results are then discussed in more detail in Section III, where the performance of the various models is compared, and in which explanations are provided for the results. Finally, Section IV wraps up the study, then offers future research directions in the area of stress detection, using ML and DL approaches.

II. METHODOLOGY

A. Dataset Description

The public “Stress Level Detection” dataset available on Kaggle includes the following features: Temperature, Step Count, Humidity, and Stress Level [23]. It is designed to assess and analyze stress levels in individuals based on their physical activity [21]. The dataset contains 2,001 samples, including body temperature, humidity, step count, and the corresponding stress level labels.

To avoid any data leaking, the stress level is treated as the target variable, so it is left out from the group of input features used during training for this study.

Although the dataset provides a useful benchmark for stress prediction, it contains only 2,001 samples, which may limit the generalizability of the developed model. Consequently, the reported performance should be interpreted within the context of this dataset, and further validation using larger and more diverse datasets is required.

B. Data Preparation

The critical factor behind this model is the quality of the data. The dataset created for this study is publicly available on kaggle and ensures an effective training, evaluating and testing using a cross sectioning approach. The method will allow a complete comparison to study and includes DL and ML models as described in the research.

Data processing workflow in fig.1 where it demonstrates the sequential steps to make a dataset ready. The process start with loading the dataset which is the first step in data manipulation. In case class imbalanced is detected, SMOTE (Synthetic Minority Over Sampling Technique) is only performed on the training data as this would lead to information leakage if it were done before train-test split. Then, handle missing values with statistical methods (mean, median or mode) or by removing incomplete records.

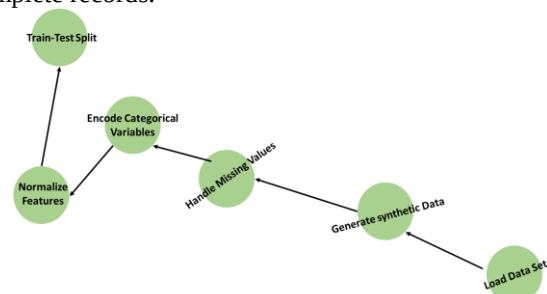


Fig. 1. Data Preprocessing workflow

After handling missing values categorical variables are converted into numerical forms so that it can be used in different models. Subsequently feature normalization is performed to keep all the features at the same level of influence. Next, it splits the data set into training and testing, to learn the model and test it. This methodical approach ensures data cleansing, boosting reliability and optimizes performance.

C. Deep Learning Model

1) Gated Recurrent Unit (GRU)

The GRU are a type of RNNs borrowed from HMMs and required for overcoming the missing sequential structure and the temporal sequencing limitations of memory cells due to the vanishing gradient problem. Unlike LSTMs, the GRUs have just two gates: an update gate and a reset gate, in contrast to three gates. The GRUs have only two gates, an update gate and a reset gate, as opposed to three gates in the LSTM model, which makes the LSTM model simpler. The GRUs have fewer parameters in their structure compared to LSTMs that help them to train faster, require less memory, easy implementation and there is still a good performance of mitigating the gradient vanishing problem. The update gate and reset gate control the number of bits of the last hidden state retained, and how many bits are added to the hidden state [19]. In tasks with longer sequences, LSTMs typically outperform, because they have a more complex memory mechanism but in cases where the sequences are relatively short and/or computational power is constrained GRUs usually beat or at least match performance of LSTMs. GRUs get used in a variety of applications including natural language processing, time series analysis, speech recognition and video analysis because they can learn to process sequential data very efficiently and effectively. Although GRUs have advantages, they may face challenges in ultra-long sequences and are less interpretable than LSTMs. In general, GRU provides a compelling and useful alternative to LSTMs whilst being particularly advantageous for mobile or resource constrained learning tasks, where training and prediction needs to occur quickly [20]. In Fig. 2: Input sequences run through a GRU-based neural network, where each unit passes information through its own dropout and dense layers, resulting in the final output. This illustrates the graph view of dataflow in one direction through a network.

Eq. (1) describes fundamental operation in RNN or other similar architecture. In here, is the output at time step t which calculate weight matrix on a new vector that merges together both the previous hidden state and the new input. If it is passed

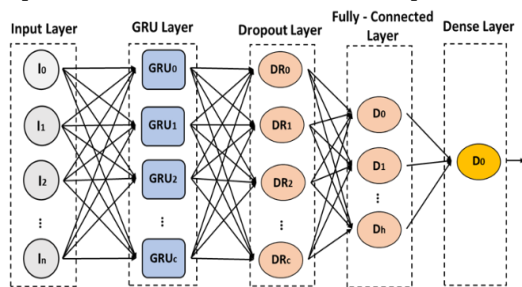


Fig.2. GRU Neural Network Structure [28]

to an activation function like sigmoid function which gives nonlinearity in output. The hidden state contains information

from earlier time steps, letting the net have kind of a memory, and is the input data at this time step. The learned parameters are inside the weight matrix which control how the inputs and embedding state affect output. Due to this, it allows a network to keep track of data that unfolds sequentially, meaning it can simultaneously update its own internal hidden state at each time step positively impacting on sequential dependencies and structural patterns of the data [28].

$$Z_t = \sigma(W_z \times [h_{t-1}, X_t]) \quad (1)$$

D. Machine learning Model

1) Adaptive Boosting (AdaBoost)

AdaBoost is a powerful ML algorithm that enhances the performance of weak classifiers by combining them into a strong ensemble model. It works by training weak classifiers sequentially, focusing on correcting errors made by previous models at each step. The algorithm assigns weights to training samples, increasing the weights of misclassified samples and decreasing those of correctly classified ones, ensuring that subsequent classifiers pay more attention to harder to classify data [29]. Once trained AdaBoost calculates a weight for each classifier based on its accuracy, and the final prediction is then a weighted vote of all classifiers. The fact that AdaBoost employs an iterative and adaptive algorithm makes it very efficient in classification tasks, as usually only a few parameters need to be set to get high classification accuracy. But AdaBoost is sensitive to noisy data and outliers as it pays too much attention to misclassifications and can overfit [35]. However, there are still some limitations to the AdaBoost algorithm [7] and they are: AdaBoost is not robust to noisy data, as it can take poor weak classifiers producing complex

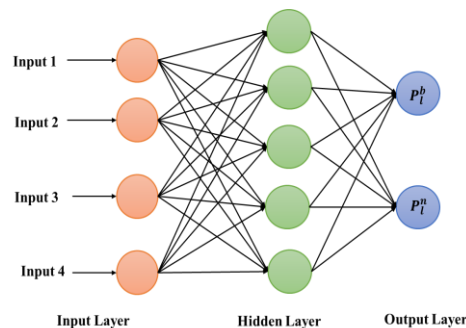


Fig. 3. AdaBoost Model Representation [36]

models; other ensemble methods prevail over it; in most of practical use situations, if used as a complete model selection process instead of for fine tuning in fig. 3. Although AdaBoost is still a very popular choice for complex classification problems, it can be seen as an efficient and effective method that leverages its strengths.

Eq. (2) The key idea of boosting algorithms such as AdaBoost (when we combine multiple weak learners into a single strong learner). Where is the weight of a weak learner at iteration t , and in most algorithms it defines how much that weak will affect the base model. Here, t indicates the error rate of the weak learner at iteration and how accurate it is when

predicting on the training data. The term reduces the weight depending on how well the weak learner performs by applying a natural logarithm function to the error ratio. when the weak learner has great accuracy then will become large, this gives it more powerful influence. But the weak learner does poorly will be small, so it has very little influence. This process works as a way of learning from the mistakes made by past weak learners to create a strong ensemble that performs very well and is resilient [36].

$$\alpha_t = \frac{1}{2} \ln \left(\frac{1-\varepsilon_t}{\varepsilon_t} \right) \quad (2)$$

2) Support Vector Machines (SVM)

The SVM is a powerful and versatile ML algorithm which can be applied to classification or regression problems in Fig. 4. It is especially designed to find the best hyperplane which can maximally separate both of these classes in our data. If the data is linearly separable, SVM generates a linear hyperplane, and if the data is not linearly separable, kernel functions are used such as polynomial, radial basis function (RBF), sigmoid etc. to transform the input features in a higher dimensional space such that linear separation is possible [33]. This versatility of being able to deal with both linear and non-linear data makes SVM really useful such as for image classification, text categorization, and bioinformatics. Nevertheless, SVM may be computationally expensive (mainly when it comes to large datasets) due to the fact that solving quadratic optimization problem [10]. In addition, it requires fine-tuning of hyperparameters, such as the choice of kernel, regularization parameter, and other kernel-specific hyperparameters. Even though these challenges SVM is the preferred one because it is robust, accurate and can out of the box generalize particularly in high dimensional space for solving complex ML problems [34].

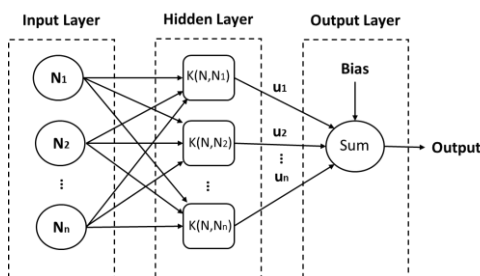


Fig.4. Architecture of a Support Vector Machine (SVM) [30]

where α_i is the Lagrange multiplier associated with the i_{th} support vector, y_i is the class label of the i_{th} training sample, $K(x_i, x)$ denotes the kernel function that measures the similarity between the support vector x_i and the input sample x and b is the bias term. The sign of the decision function determines the predicted class label.

$$f(x) = \text{sign}(\sum_{i=1}^n \alpha_i y_i K(x_i, x) + b) \quad (3)$$

3) Light Gradient Boosting Machine (LightGBM)

LightGBM, or Light Gradient Boosting Machine, is a fast and scalable gradient boosting framework that uses the histogram-based learning algorithm to achieve high efficiency on very large datasets for faster training while maintaining accuracy. LightGBM takes a different approach than other gradient-boosted decision trees, as it uses leaf-wise growth, meaning the tree will split on the leaf that has yielded the greatest loss gain. This way, they reduce the training time and give an often higher accuracy since it only aims for the most important splits. In addition to being fast and super scalable on clusters of computers, LightGBM is also able to process categorical features without converting them to one-hot-encodings [35] that makes it particularly conducive to large-scale data. It can also effectively manage missing data, learning how to optimally handle null values during training in Fig. 5. LightGBM, on the other hand, is good for big data sets, but can over-fit in small data sets if not tuned properly. The most important hyperparameters such as number of leaves, learning rate and regularization terms need to be tuned in order to balance model complexity with generalization. In each of these applications from ranking to classification to regression, however, LightGBM is arguably one of the most widely used algorithms and the speed, scalability and performance benefits continue to keep LightGBM popular with data scientists and ML practitioners.

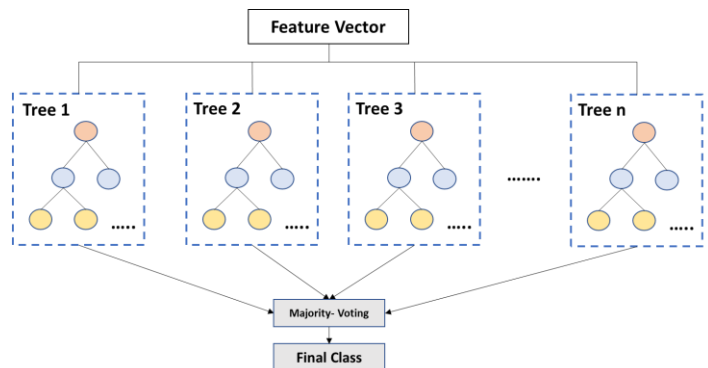


Fig. 5. LightGBM Algorithm architecture [37]

where L is the overall objective function to be minimized, $l(y_i, \hat{y}_i)$ is the loss function that measures the difference between the actual value y_i and the predicted value $\hat{y}_i$, λ is the regularization parameter controlling model complexity, $\Omega(T)$ represents the regularization term associated with the complexity of the decision tree T , and T denotes the tree structure. The objective function balances prediction accuracy and model complexity to improve generalization.

$$L = \sum_{i=1}^N l(y_i, \hat{y}_i) + \lambda \Omega(T) \quad (4)$$

Fig. 6 provides a framework you can follow to create a ML model based on health data. Start of the process is data collection with some tracking metrics like step counting, humidity and body temperature. They are kept as separate dataset because they have training, validation and test sample. Model tuning Hyperparameter optimization for the model tune it which is like AdaBoost and SVM and LightGBM. It includes

an evaluation of the final classifier (the model) to validate that its predictions are reliable and accurate for predicting health outcomes. A systematic approach guarantees sealed model development and validation [36].

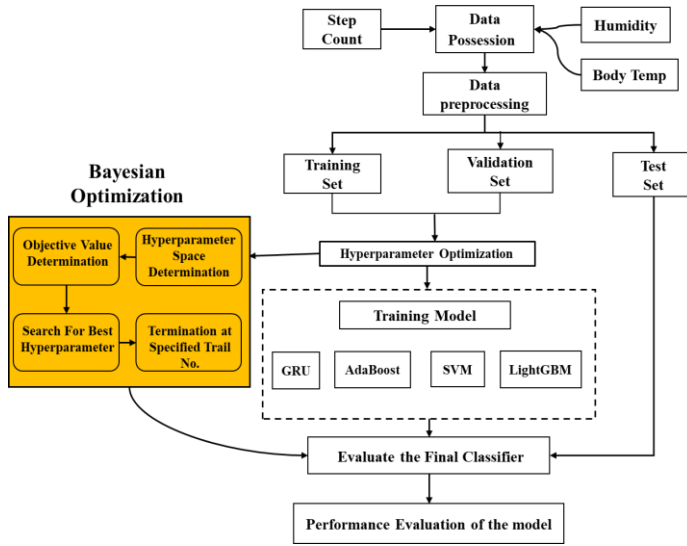


Fig.6. Proposed flowchart to detect the model

E. Bayesian Hyper Parameter Optimization

Hyperparameter tuning is a critical step within ML which improves model performance by hyper tuning the parameters governing a learning algorithm. Eq. And minimization of (5) are optimization problems in finding the better hyperparameter on the separate test set with respect to the loss function L . The training set is used to obtain using the model C , and will then be iteratively modified in order to improve performance with respect (w.r.t. This modelling is mainly driven from new data so that the train and test are balanced giving the effectiveness Hyper Parameter tuning in mathematical terms, this is written as:

$$\lambda^* = \arg \min_{\lambda} L(X_{TS}; C(X_{Tr}; \lambda)) = \arg \min_{\lambda} F(X_{TS}, X_{Tr}, C, \lambda, L) \quad (5)$$

The traditional hyperparameter tuning methods, such as grid search and random search, are based on an exhaustive or random sampling of the space that can be computationally expensive and time-consuming. Bayesian optimization is not a random search of random combinations of parameters, but rather it narrows the search space by using prior knowledge and evaluating only the best in-class model in successive iterations [37].

By modeling the performance of parameter combinations of the hyperparameters, Bayesian optimization performs a relatively small number of experiments to find an optimum solution, while traversing the search space. The hyper parameters tune for selected ML classifiers and hyper parameter search space are summarized in Table I, which gives a systematic foundation to tune and increase the performance of them.

TABLE I. HYPERPARAMETER SEARCH SPACE FOR DIFFERENT MODELS

Model	Selected Hyperparameter	Type	Search Space
LightGBM	n_estimators	Integer	[50 ,200]
	learning_rate	Continuous	[0.001 - 0.3]
	max_depth	Integer	[3 ,15]
	subsample	Continuous	[0.5, 1]
	colsample_bytree	Continuous	[0.5, 1]
GRU	units	Integer	[32, 256]
	learning_rate	Continuous	[0.0001, 0.01]
	dropout	Continuous	[0.2, 0.5]
	batch_size	Integer	[16, 128]
	epochs	Integer	[10, 100]
AdaBoost	n_estimators	Integer	[50, 200]
	learning_rate	Continuous	[0.01, 2.0]
	algorithm	Categorical	[SAMME, SAMME.R]
SVM	kernel	Categorical	[linear, poly, rbf, sigmoid]
	C	Continuous	[0.1,100]
	gamma	Categorical	[scale, auto]

F. Performance Evaluation Metrics

Four main metrics are used to evaluate the effectiveness of ML and DL models, True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). In this con-text, a positive instance corresponds to a stress predicted whereas a negative instance refers to as a "normal state" (no stress predicted). These are simple metrics that can be used to compute the most critical demonstrate performance measures such as accuracy, precision, recall and F1-score, which can be used to assess system performance over a range of classification tasks [38].

True Positive (TP): Stress condition actually is present in data, and the algorithm says that this is a stress.

True Negative (TN): When the stress is not present in the data, and the classifier correctly labels it as not occurring.

False Positive (FP): A true condition of stress is indicated in the actual data, but the algorithm indicates it to be a failure.

False Negative (FN): Actual condition is stress; however, algorithm does not detect it in the data.

One of the key metrics used to assess the performance of a classification system is called accuracy(Acc). It is de-fined as the proportion of correct classification (both in true positive and true negative) to overall sample size.

$$Acc = \frac{TP+TN}{Total\ Predictions} \quad (5)$$

Precision (P) measures the proportion of correctly predicted stress-positive instances among all instances predicted as stress. It is defined as:

$$P = \frac{TP}{TP+FP} \quad (6)$$

Recall (R) or sensitivity, refers to the true positive rate. It measures the sensitivity of the system if there are actually the conditions for stress. Recall equals:

$$R = \frac{TP}{TP+FN} \quad (7)$$

The F1-score is a metric that combines precision and recall into a single value, providing a balanced measure of a model's

performance. It is the harmonic mean of precision (P) and recall (R), particularly useful when there is an imbalance between classes. The F1-score is defined as:

$$\text{F1 Score} = 2 \times \frac{P \times R}{P + R} \quad (8)$$

III. RESULTS

The stress level is considered as the dependent variable or the desired dependent variable in this study, or for prediction, and temperature, humidity and steps are also the input variables to the prediction. Based on the experimental results, these contextual features fed to the model seem to be of relatively low-dimensional space but still contain information meaningfully related to the stress condition under which it is considered.

An appraisal of the ML and DL models was conducted using Python, in a Jupyter Notebook, and some variations in the effectiveness of the stress detection process were observed. Some of the models were more accurate and robust after they were subjected to some kind of Bayesian hyperparameter tuning or near it, and with a better generalization. The second point was that the graphs were helpful as well sometimes not from accuracy, precision, recall and the F1 score, but the other graphs such as confusion matrices, precision recall curves etc. that provided a certain context. Despite some models developing errors in classification, others were quite consistent and reliable, and generally, it provides a sense that they could be useful in a light-weight use of stress prediction.

Fig. 7 shows how accurate the different ML models are. LightGBM ends up with the best accuracy around 0.97, then GRU is next at about 0.91. There's SVM at 0.86 after that. Finally, AdaBoost has the smallest accuracy around 0.55 and that one is definitely the lowest.

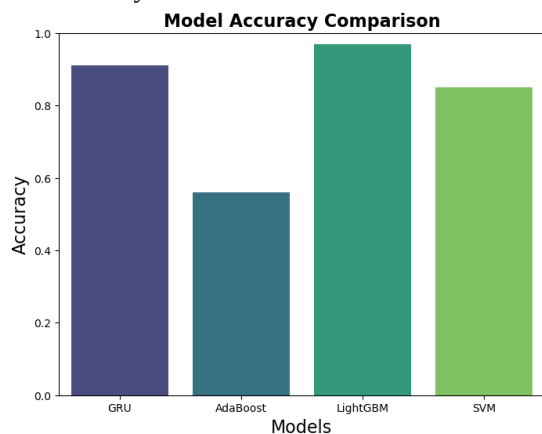


Fig.7. Model Accuracy comparison

The Fig. 8 graph shows the comparison of Precision, Recall and F1-Score of four ML models. The scores are led by LightGBM (0.95, 0.95, 0.95) followed closely by GRU (0.92, 0.90, 0.91) and SVM (0.85, 0.87, 0.86). The AdaBoost performs the worst with a scores of 0.60, 0.65 and 0.62 respectively. The performance of AdaBoost is by-land behind LightGBM.

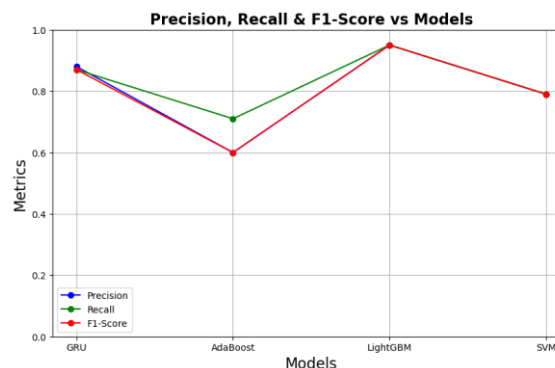


Fig.8. Based on the models performance comparison

For a few distinct models (GRU, AdaBoost, LightGBM and SVM), the accuracy, precision, recall and F1-score is shown in Table II, which is a summary of their accuracy and precision in classifying the dataset we used. In particular, LightGBM appears to perform best on basically all metrics, with the best results on all. In the meantime, AdaBoost becomes the weakest among all, with the least performance. As can be seen in Fig. 9, a high accuracy is achieved by LightGBM with few misclassifications in the confusion matrix. As most of the predictions, it shows that the errors are well very small, indicating high precision and reliability in the prediction of stress.

TABLE II. COMPARATIVE PERMANCE MATRICS (WITH AND WITHOUT OPTIMIZATION) FOR STRESS DETECTION

Models	GRU	AdaBoost	LightGBM	SVM
Accuracy	0.88	0.52	0.93	0.80
Accuracy (with Optimization)	0.91	0.56	0.97	0.85
Precision	0.85	0.55	0.91	0.74
Precision (with Optimization)	0.88	0.60	0.95	0.79
Recall	0.84	0.65	0.91	0.74
Recall (with Optimization)	0.87	0.71	0.95	0.79
F1 Score	0.84	0.55	0.91	0.74
F1 Score (with Optimization)	0.87	0.60	0.95	0.79

The ROC curve in Fig.10 kind of shows the classification performance of LightGBM, AdaBoost, SVM and GRU based on their AUC (Area Under Curve) values. LightGBM ends up with the top AUC at 1.00, then come GRU (0.991) and SVM (0.991) almost at the same level. AdaBoost is at the bottom, at about 0.980, suggesting the overall performance is a bit suboptimal. In general, LightGBM outperformed the other models evaluated, such as AUC of 1.00 in the experimental dataset.

Although the proposed model achieved high classification performance, these results should be interpreted with caution because the evaluation was conducted on a relatively small public dataset. Train-test separation and Bayesian hyperparameter optimization were employed to reduce overfitting; however, additional validation using larger and independent datasets is necessary to confirm the robustness and generalizability of the proposed framework.

The reported performance metrics were obtained using a single train–test split of the dataset. Although Bayesian hyperparameter optimization was employed to improve model performance and reduce overfitting, the results may still be influenced by the limited size of the dataset. Hence, further testing through replicated experiments and k-fold cross-validation is suggested to validate the robustness and stability of the proposed framework.

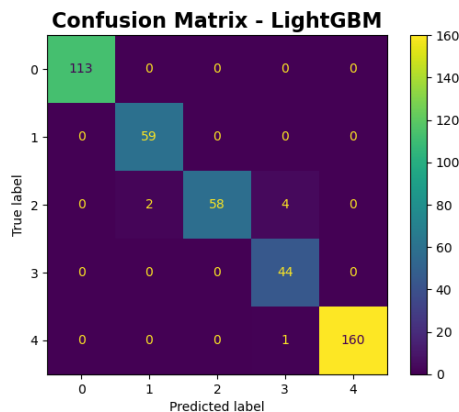


Fig. 9. Confusion matrix for the proposed model

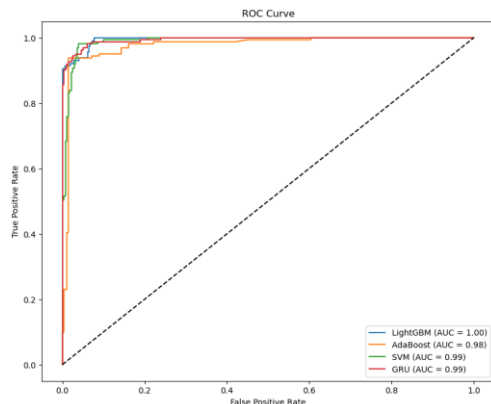


Fig.10. Compare classification performance of different models

Table III gives kind of a comparative view of several ML and DL models for stress detection, based on the main performance measures like accuracy, precision recall and F1-score. Of them LightGBM seems to be the best with an accuracy of 0.97, precision of 0.95, recall of 0.95 and F1 score of 0.95, which really demonstrates its ability to handle and learn from the data set. Random forest also performs quite well with accuracy 0.92 and f1 score 0.93, so it is also a reliable alternative. Similarly, KNN + SVM works well, having accuracy of approximately 0.90 and recall of 0.91 indicating that it is quite effective in processing the dataset. Meanwhile CNN achieves good accuracy of 0.89 and recall of 0.90; this shows that CNN is able to capture the structure of the data. However, CNN + BiLSTM and DNN on the other hand, have kind of moderate performance, both having an accuracy of 0.68 and 0.78 respectively, which kind of indicates that they are not very good at tackling the complexity of the dataset, or maybe they're not.

TABLE III. PERFORMANCE COMPARISON WITH EXISTING STRESS PREDICTION APPROACHES.

Model	Accuracy	Precision	Recall	F ₁ Score
KNN+SVM [22]	0.90	0.89	0.91	0.88
CNN+BiLSTM [27]	0.68	0.65	0.70	0.70
DNN [18]	0.78	0.79	0.78	0.80
CNN [21]	0.89	0.85	0.90	0.85
SVM+AdaBoost [18]	0.56	0.60	0.60	0.65
SPFHD [30]	0.87	0.85	0.84	0.79
Random Forest [18]	0.92	0.91	0.90	0.93
LightGBM (Proposed Model)	0.97	0.95	0.95	0.95

However, SVM + AdaBoost appears to underperform significantly with low accuracy (0.56) and low F1 score (0.65) and the combination Performance Comparison with other Stress prediction Approaches is not quite as smooth. Overall, SPFHD is a good or a decent model as it gets an accuracy of 0.87 and an F1 score of 0.79 meanwhile. Overall, the findings suggest that LightGBM and Random Forest models perform better in stress detection tasks, while also revealing certain limitations in other models, like SVM + AdaBoost.

IV. CONCLUSION

The goal of this study is to predict stress by considering feature sets based on environmental and physical activity parameters, such as temperature, humidity, distance travelled (in steps), and using a machine-learning approach. The model proposed using Light GBM with Bayesian hyper parameter optimization gave a better result compared to Random Forest where accuracy was 97%, precision and recall was 95% each and F1 score was also 95%. The results were superior compared to other tested models like a GRU, an SVM and an AdaBoost with an accuracy of 91%, 86% and 55% respectively. Additionally, a few models that were feature-specific, in conjunction with the reasoning behind temperature and humidity being important predictors aided the interpretability and reliability of the system. The proposed framework is very effective for the tested data-set as it achieves an AUC of 1.00 and high classification accuracy. However, for real world deployment it still needs to have larger and more varied data sets, and data needs to be collected independently. Also, the experimental material was evaluated using only one train–test split. Future studies should include k-fold cross validation and repeated experimental trials to obtain a statistically valid estimate of the performance of the proposed model and to validate the consistency of the proposed model.

The proposed framework will be validated using larger and more representative data sets collected from diverse populations in the future. This validation will aid in the evaluation of the robustness of the proposed stress prediction model, its generalizability, and its applicability to real-world scenarios.

REFERENCES

- [1] Abd Al-Alim, M., Mubarak, R., Salem, N. M., & Sadek, I. (2024). A machine-learning approach for stress detection using wearable sensors in free-living environments. *Computers in Biology and Medicine*, 179, Article 108918.
- [2] Abdelfattah, E., Joshi, S., & Tiwari, S. (2025). Machine and deep learning models for stress detection using multimodal physiological data. *IEEE Access*, 13, 4597–4608.
- [3] Xiang, J.-Z., Wang, Q.-Y., Fang, Z.-B., Esquivel, J. A., & Su, Z.-X. (2025). A multi-modal deep learning approach for stress detection using physiological signals: Integrating time and frequency domain features. *Frontiers in Physiology*, 16, Article 1584299.
- [4] J. A. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 2, pp. 156–166, Jun. 2025.
- [5] Dev, U., Ahmad, S., Uddin, M. R., Kabiruzzaman, M., Dev, S., Hazari, M. R., & Hossain, C. A. (2026). Seasonal and altitudinal wind energy potential assessment at varying heights using stochastic modeling approaches. *AIUB Journal of Science and Engineering*, 24(3), 238.
- [6] A. Sano and R. W. Picard, "Stress recognition using wearable sensors and mobile phones," in *Proc. 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Osaka, Japan, 2019, pp. 6737–6740.
- [7] Y. Can, B. Amrich, and C. Ersoy, "Stress detection in daily life scenarios using smartphones and wearable sensors: A survey," *Journal of Biomedical Informatics*, vol. 92, Art. no. 103139, Apr. 2019.
- [8] P. Schmidt, R. Reiss, R. Dürichen, C. Marberger, and K. Van Laerhoven, "Introducing WESAD, a multimodal dataset for wearable stress and affect detection," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 3, Art. no. 112, Sep. 2018.
- [9] A. Alberdi, A. Aztiria, and A. Basarab, "Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review," *Journal of Biomedical Informatics*, vol. 59, pp. 49–75, Feb. 2019.
- [10] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas, "Taking the Human Out of the Loop: A Review of Bayesian Optimization," *Proceedings of the IEEE*, vol. 104, no. 1, pp. 148–175, Jan. 2020.
- [11] Singh, A., Gupta, R., & Kumar, V., "CNN-based physiological signal analysis for automated stress detection," *IEEE Access*, vol. 9, pp. 45763–45775, 2023.
- [12] Fernandes, J., Silva, H., & Costa, P., "Blood pressure and galvanic skin response as key indicators of stress: A machine learning approach," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 5, pp. 1234–1242, 2022.
- [13] Alberdi, A., Aztiria, A., & Basarab, A., "Early stress recognition in office environments using multimodal physiological data and deep learning," *Sensors*, vol. 21, no. 3, p. 678, 2020.
- [14] Plunkke, J., Schneider, M., & Bauer, G., "Stress classification in firefighters using heart rate variability and machine learning," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 7, pp. 1890–1902, 2020.
- [15] Sano, A., & Picard, R. W., "Stress recognition using wearable sensors and mobile phones," *Proceedings of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 4744–4747, 2021.
- [16] Sandulescu, A., Rodríguez, C., & Fernandez, J., "Fuzzy logic-based stress detection system leveraging deep learning models," *Applied Soft Computing*, vol. 134, p. 109876, 2019.
- [17] Ratul, M. H. A., Tasnim, S., Durjoy, M. M. A. P., Shufian, A., Uddin, M. R., & Hossain, M. S. (2026). NetZeroML: Hybrid PCA-XGBoost framework for global decarbonization and renewable energy forecasting. In *2026 IEEE Region 10 Symposium (TENSYP 2026)*.
- [18] T. Elzeiny et al., "Automatic stress detection using voice and physiological signals," *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3672–3676, 2022.
- [19] H. R. Fajrin, Sasmeri, L. R. Prilia, B. Untara, and M. A. F. Nurkholid, "Fuzzy logic method-based stress detector with blood pressure and body temperature parameters," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 14, no. 2, pp. 2156–2166, Apr. 2024.
- [20] T. Periyaswamy and M. Balasubramanian, "Combining multiple human physiological signals using fuzzy logic to determine stress caused by battle dress uniforms," *Discover Applied Sciences*, vol. 4, Art. no. 319, 2022.
- [21] Y. Zhang, L. Wang, and J. Liu, "Facial expression analysis for stress detection using deep learning," *IEEE Transactions on Affective Computing*, vol. 11, no. 4, pp. 567–579, 2020.
- [22] H. Gjoreski, M. Gjoreski, and M. Gams, "Context-aware stress detection using wearable sensors and smartphones," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 8, pp. 3121–3132, 2021.
- [23] Kaggle, "Stress Level Detection." Available online: <https://www.kaggle.com/datasets/laavanya/stress-level-detection> (accessed Dec. 15, 2024).
- [24] J. Smith, A. Brown, and P. Taylor, "Explainable artificial intelligence for physiological stress detection," *IEEE Transactions on Human-Machine Systems*, vol. 52, no. 3, pp. 456–467, 2022.
- [25] J. Perez, R. Martinez, and L. Garcia, "Federated learning for stress detection: A privacy-preserving approach," *IEEE Internet of Things Journal*, vol. 8, no. 12, pp. 9876–9887, 2021.
- [26] L. Wang, Y. Zhang, and X. Liu, "Environmental factors and stress detection using IoT sensors and physiological data," *IEEE Sensors Journal*, vol. 23, no. 5, pp. 5678–5690, 2023.
- [27] A. Kumar, S. Sharma, and R. Singh, "Thermal imaging-based stress detection using deep learning," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–12, 2022.
- [28] Dev, U., Ahmad, S., Uddin, M. R., Dhar, R., Mubarak, H., Hazari, M. R., & Mekhilef, S. (2025). Short-term wind speed forecasting using different LSTM deep learning methods. In *Innovations in Electrical and Electronics Engineering (Lecture Notes in Electrical Engineering, Vol. 1295)*, pp. 199–211. Singapore: Springer Nature.
- [29] M. Hernandez, L. Gomez, and P. Sanchez, "Social media activity analysis for stress detection using NLP techniques," *IEEE Transactions on Computational Social Systems*, vol. 9, no. 2, pp. 345–356, 2022.
- [30] T. Nguyen, H. Tran, and Q. Le, "Multimodal data fusion for stress detection using deep learning," *IEEE Transactions on Affective Computing*, vol. 14, no. 1, pp. 123–135, 2023.
- [31] R. Gupta, S. Kumar, and A. Jain, "Gait analysis for stress detection using machine learning," *IEEE Sensors Journal*, vol. 21, no. 15, pp. 16789–16800, 2021.
- [32] J. Lee, K. Kim, and S. Park, "Transfer learning for stress detection in small datasets," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 5, pp. 2345–2356, 2023.
- [33] X. Chen, Y. Li, and Z. Wang, "Voice-based stress detection using deep learning models," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 2150–2161, 2021.
- [34] H. Kurniawan, J. Maslovsky, and M. Pfurtscheller, "EEG-based stress detection using machine learning techniques," *Frontiers in Neuroscience*, vol. 7, no. 1, pp. 1–9, 2023.
- [35] Dev, U., Uddin, M. R., Ahmad, S., & Islam, A. (2026). Edge cutting and AI-driven protection strategies for DC microgrids: A comprehensive survey of challenges, technologies, and future trends. *Results in Engineering*, 30, Article 110013.
- [36] M. T. Islam et al., "Ensemble learning-based stress classification using wearable sensor data," *IEEE Sensors Journal*, vol. 20, no. 17, pp. 10215–10227, 2020.
- [37] Uddin, M. R., Zishan, M. S. R., Ahmad, S., & Dev, U. (2026). Enhancing stress detection using Bayesian optimized boosting algorithms and feature analysis using SHAP. In *ICT for Intelligent Systems (ICTIS 2025) (Lecture Notes in Networks and Systems, Vol. 1510)*. Singapore: Springer.
- [38] Dev, U., Ahmad, S., Uddin, M. R., Hazari, M. R., Kabiruzzaman, M., Zishan, M. S. R., & Ahmed, K. F. (2025). Day ahead wind energy forecasting in two distinct locations of Bangladesh using deep learning models. In *2025 International Conference on Electrical, Computer and Communication Engineering (ECCE)* (pp. 1–6).



Mohammad Rabib Uddin has received his B.Sc. degree in Electrical and Electronic Engineering (EEE) from American International University-Bangladesh (AIUB) and received his M.Sc. in EEE from AIUB. He has worked as Teaching Assistant in the faculty of engineering at AIUB. He is currently working as a Lecturer in the faculty of engineering at AIUB. The areas of his research focus include Artificial Intelligence, VLSI, Power Electronics and Electronic Devices. He can be contacted at email: rabib@aiub.edu

learning. She can be contacted at email: 23-93050-2@student.aiub.edu



Md. Saniat Rahman Zishan received his B.Sc. degree in Electrical and Electronic Engineering and his Master of Engineering degree in Telecommunications from the American International University-Bangladesh (AIUB). He completed his Ph.D. in the field of e-Health at Universiti Sultan Zainal Abidin. He began his teaching career as a Lecturer at AIUB in September 2009 and is currently serving as a Professor and Director of the Faculty of Engineering. He also held the position of Special Assistant in the Office of Student Affairs (OSA) from December 2012 to September 2018. Mr. Zishan has published over 50 articles in various journals and at international and national conferences. His research interests include e-Health, wireless communication, signal processing, telemedicine, and robotics. He is a member of the Institute of Electrical and Electronics Engineers (IEEE) and the Institution of Engineers, Bangladesh (IEB).



Shameem Ahmad has received his Ph.D. degree in electrical engineering from Universiti Malaya, Malaysia, in 2022. He is currently serving as an Associate Professor in the Department of EEE, BSRM school of Engineering, BRAC University. His research interests include microgrid, smart grid, power system control, power converter control and electric vehicle.



Upama Dev received her B.Sc. degree in Electrical and Electronic Engineering and her M.Sc. degree in EEE, both from the American International University-Bangladesh (AIUB). Her research interests include electric vehicles, smart grids, renewable energy, and machine