

Received 7 September 2023, accepted 9 November 2023, date of publication 14 November 2023,
date of current version 20 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3332762

RESEARCH ARTICLE

PredictEYE: Personalized Time Series Model for Mental State Prediction Using Eye Tracking

C. JYOTSNA^{1,3}, J. AMUDHA¹, AMRITANSHU RAM², DAMIANO FRUET³,
AND GIANDOMENICO NOLLO³, (Member, IEEE)

¹Department of Computer Science and Engineering, Amrita School of Computing, Amrita Vishwa Vidyapeetham, Bengaluru 560035, India

²HCG Cancer Center, Bengaluru 560027, India

³Department of Industrial Engineering, University of Trento, 38123 Trento, Italy

Corresponding author: C. Jyotsna (j.chandrasekharan@unitn.it)

This work was supported by the Erasmus International Credit Mobility, which aided in developing the project ideas and the model, under Project 2019-1-IT02-KA107-061375.

This work involved human subjects in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Committee of the HCG Cancer Hospital, Bengaluru.

ABSTRACT Mental health is vital for emotional, psychological, and social well-being. Mental illness can affect thoughts, feelings, and behaviors. Early intervention and specialized care can manage major mental illnesses. Predicting mental state accurately can facilitate behavioral changes and promote overall well-being. The paper proposes a novel personalized time series model called PredictEYE, which aims to predict a person's mental state and identify the specific scene responsible for that mental state. The model achieves this by analyzing individuals' eye-tracking time series data while watching calm and stressful videos. The model utilizes a deep learning time-series univariate regression model based on Long Short-Term Memory for predicting the future sequence of each feature and a machine learning-based Random Forest algorithm for the mental state prediction. The model's performance was compared across the state-of-the-art literature survey. The predictEYE model could achieve an accuracy of 86.4% accuracy in predicting mental state. Tailoring eye tracking models to individual differences is more effective in comprehending mental states than models that make comparisons across multiple participants, given eye tracking data's unique and distinctive idiosyncratic nature. The eye tracking features play a crucial role in predicting the mental state, and the model is adaptable to work with webcam-based eye tracking and can relate to applications where continuous and non-invasive monitoring is required.

INDEX TERMS Eye tracking, galvanic skin response, long short-term memory, mental state prediction, time series analysis.

I. INTRODUCTION

The growth, development, and productivity of society depend on health, which is essential and significant for a happy and healthy life anywhere in the world. Mental health is an integral and essential component of health. Everyone can be affected by mental disorders, irrespective of age and gender. Brain and mind-related health conditions are increasing daily, and it has become necessary to monitor mental health regularly. So everyone in society requires accurate measures

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

for mental health monitoring [1], [2]. Mental health and physical health are closely linked. Mental illness, such as depression and anxiety, affects our ability to participate in healthy behavior. It plays a vital role in maintaining good health [3], [4].

In many countries around the world, primary healthcare systems typically consist of a combination of community health centers, primary health clinics, district or regional hospitals, and other specialized institutions. Integrating mental health care into primary care programs is a way to treat mental health conditions in the primary care system [5]. It can involve treating a person entirely with

existing health problems. However, as with almost all medical issues, it carries a stigma, and those with mental illness become even more ineffective in the workplace. Growing evidence indicates that receiving mental health care faces much more challenging obstacles than receiving physical care [6]. The shortage of mental health professionals leads to the unavailability of needed care. Financial constraint is another mental health service issue that prevents people from accessing necessary therapy. Therefore, low-cost solutions to indicate mental illness early in its development are essential.

One of the primary reasons for mental illness is stress. Stress in an individual activates the sympathetic nervous system and releases hormones that cause changes throughout the body. Stress induces many physiological changes in the body. It causes an increase in blood pressure, heart rate, sweating, pulse, pupil size, quicker breathing, muscle tension, dry mouth, and accelerated heart rate [7]. A person's stress level is measured based on their mental workload, cognitive ability, and emotional or mental state. A person's mental state can vary based on physiological and cognitive conditions. There are mental states like calm, stress, drowsiness, distraction, cognitive load, and mind wandering [8].

A person's mental health can be assessed subjectively or objectively [9], [10], [11]. Subjective measurements refer to information based on personal opinions, feelings, and perceptions. It can be measured based on well-tested, standardized questionnaires that anyone can undertake to assess their mental health. These measurements have two serious shortcomings: they can't be taken in real-time, and they are subject to psychological bias. Objective measurements refer to information that is based on observable and measurable facts. Personal biases or opinions do not influence it. Objective measurements can be obtained based on physiological and behavioral measures. Audio video recordings are used to collect behavioral data. Physiological measurements are frequently utilized to monitor mental health because they allow for unobtrusive and involuntary data collection [12]. The physiological aspects of a patient, including their eye movements, brain activity, heart rate, dermal activity, facial expressions, voice modulations, temperature, respiratory rate(BPM), and blood pressure, are sensed by various technologies to help track mental health (BP) [13].

Remote [14], [15] and personalized [16], [17], [18] monitoring are two approaches to track health and wellness data that have gained significant attention recently. Remote monitoring involves technology to monitor an individual's health and wellness remotely. Remote monitoring in the healthcare sector could empower physicians to deliver high-quality care by keeping patients safe and healthy [15]. With multiple sensing systems, physicians can track patients' health effectively, monitor remotely, and provide immediate care [15].

Personalized monitoring involves tracking health and wellness data specific to an individual. It includes tracking dietary intake, sleep patterns, and other vital parameters that

can impact health [18], [19]. Individuals can learn more about their health and wellness by keeping track of this data. They can use this information to make wise decisions about changing their lifestyles in ways that can enhance their general welfare. On the other hand, A non-personalized model is trained on a larger dataset that is not specific to any individual or group. A non-personalized model aims to provide recommendations or generic predictions that are applicable to a larger population.

Developing a personalized model involves data collection, feature engineering, model selection, training, validation, deployment, and monitoring. Relevant features are identified, and an appropriate model is chosen for the research question. The model is trained, validated, deployed, and continuously monitored and updated for optimal performance in a real-world setting.

Many researchers have built systems based on machine learning [20], [21], [22], [23], [24], [25], and deep learning [16], [20], [26], [27], [28], [29] approaches. The deep learning-based time series personalized model has the advantage of capturing patterns and changes over time, leading to more accurate and personalized predictions. Here we propose a novel predictEYE model that utilizes a deep learning-based time series model to accurately predict a person's mental state and identify the specific scene in a video that triggers that mental state [30]. PredictEYE forecasts the mental state based on the eye-tracking data obtained while watching calm and stressful videos. Furthermore, our model goes beyond just detecting the mental state; it can also pinpoint which content in the video induces that particular mental state. This unique feature distinguishes our model from other physiological measures-based models typically used to assess mental states. These findings are presented in this technical paper.

The paper makes a significant contribution to the field of time-series prediction and mental state assessment. We have developed a novel framework called PredictEYE, which combines the power of univariate Long Short-Term Memory (LSTM) and the Random Forest algorithm to forecast future eye gaze data sequences and predict an individual's mental state based on these forecasts [31]. It also aims to identify the scene responsible for that mental state. By utilizing time-series eye gaze data, our model can accurately predict future sequences, offering valuable insights into how individuals' visual attention might evolve over time. To validate the effectiveness of PredictEYE, we incorporate the user's Galvanic Skin Response (GSR) data, which serves as a benchmark for assessing the model's performance. Comparison with GSR data demonstrates the high efficacy of our approach, highlighting its potential in accurately capturing and predicting mental states.

The paper is organized as follows. A study on related works was performed and reported in Section II. A description of the methodology adopted in PredictEYE is provided in Section III. The implementation specifics for forecasting

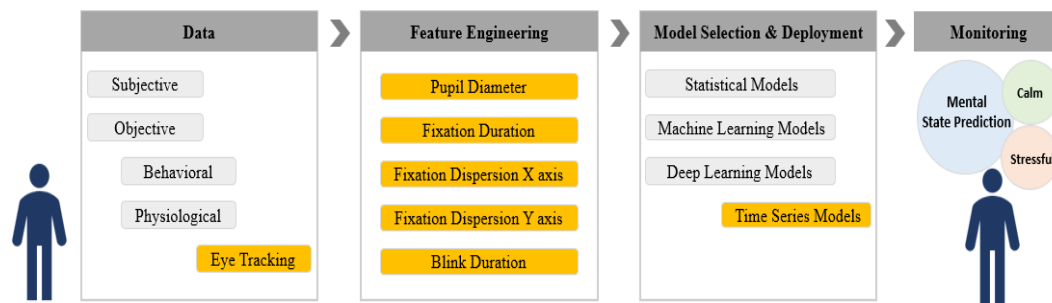


FIGURE 1. Personalized model.

the user's mental state are provided in Section IV. The outcomes of the model are discussed in Section V. Benefits of the model compared to other models are explained in Section VI. Section VII concludes by presenting the findings of this research study and highlighting potential future improvements.

II. RELATED WORKS

Mental well-being is dealing with life and its various pressures and challenges. Along with the social determinants of health, other factors can increase our stress levels and harm our sense of well-being [32]. Developing a personalized model for monitoring mental health requires careful consideration of data collection, feature engineering, model selection, model deployment, and monitoring, as shown in Figure 1. The process can be complex but can improve mental health outcomes and provide personalized support to needy individuals.

To develop a personalized model 1 for monitoring mental health, researchers and practitioners need to explore different types of data that can inform mental health statuses, such as physiological signals, behavioral data, self-reports, social media [33], [34], or environmental factors. They also need to consider different models or approaches to leverage these data to predict, classify, or diagnose mental health issues, such as machine learning, deep learning, signal processing, or natural language processing. Based on this, the literature survey is performed on various types of data used and models available for building such a model.

A. DATA AND FEATURE ENGINEERING

The data collected for mental health monitoring can be categorized as either subjective or objective, with subjective data being based on personal experiences and perceptions and objective data being based on observable, measurable factors. Subjective data are mostly collected using questionnaires, mock interviews, polling, and checklists to assess mental wellness. Standardized and well-tested questionnaires like NASA Task load Index (TLX) [9], [12], [35], Trier Social Stress Test [10], State-Trait Anxiety Inventory (STAI) [11], [16], Karolinska Sleepiness Scale (KSS) [9], [11], Shortened State Stress Questionnaire (SSSQ) [11] Stanford sleepiness Scale (SSS) [9], Wardwick – Edinburg Mental Wellbeing

Scale (WEMWBS) [16] are used for the assessment. The main limitations of the subjective data are that they can not be collected in real-time, and there can be psychological biases. Carelessness and memory lapse can lead to wrong assessments [36].

The objective data are usually collected using non-invasive wearable devices or audio-video recordings, making the data collection easier. The objective data include behavioral and physiological data. The behavioral data, facial expressions [37], audio signals [38], gestures [39], head [40], [41], hand [42], leg [43], eye movements [44] and eye contacts [45], [46] are collected using audio and video recordings or using wearable devices. The behavioral data can be collected in real-time using any non-invasive devices. Microphones or acoustic sensors are used to collect the audio signals. The speech and voice modulations are analyzed to detect mental illness [47], [48]. Digital cameras and webcams can detect people's movements and extract mental health-related data. The typing speed, mouse usage speed, the texts used in messages and emails, body postures, and time spent with mobile and computer are considered the behavioral data for the analysis [49]. The limitation of behavioral data is there can be voluntary control of the gestures, and their expressions can vary according to the culture and language. Hence, the assessments can go wrong.

Physiological signals help an individual's real-time, continuous, and consistent monitoring. The physiological signals generated in the body are involuntary and won't vary according to the culture or language. Therefore, it can indicate mental illness more efficiently than other modalities. Various sensing technologies like eye tracking [30], [50], [51], [52], accelerometers, gyroscopes, pupil corneal reflection, head tracker, Electroencephalogram (EEG) [35], [53], Electrocardiogram (ECG), photoplethysmogram, GSR, pressure-sensitive touch screen [54], [55], heart rate sensor, sound sensor, temperature sensor, oxygen saturation sensor, salivary cortisol detector [56] etc. are used for monitoring mental health [13].

Here, we mainly focus on physiological data obtained from our eyes. Eye tracking is a sensor technology that lets one know where the eyes are focused. It helps to find where people look and what they see. A strong connection exists between attention and eye movements based on the

“eye-mind hypothesis” [57]. Eye tracking technology helps tap into subconscious processing and identify the elements that attract immediate attention. The aspects that attract above-average attention, the parts ignored, and the order in which the details are noticed can be identified with the help of eye tracking [58]. Eye tracking technology provides insights into participants’ eye movements while engaging in various activities and can illuminate the unconscious mechanisms that underlie human behavior. Eye Tracking helps to measure cognitive load, concentration, focus, drowsiness, consciousness, and other mental states [50], [59]. Technologies to track the eye have become efficient, cheap, and compact and are increasing use in many fields, including gaming, driver safety, military, education [60], product recommendation system, psychology research, cognitive studies [51], market analysis, medical research [52], [61], [62], advertising [63], and healthcare [64], [65], [66], [67]. Eye tracking can provide valuable information about emotional responses, making it an important tool in mental health monitoring.

Eye tracking data obtained for each person can be considered idiosyncratic, which means that it is unique and specific to that individual [68]. The way a person gazes, fixates, and responds to different stimuli or mental states can vary significantly from one person to another [69]. This uniqueness arises due to a combination of factors, such as individual differences in cognitive processing, attentional focus, emotional responses, and eye movement patterns.

The galvanic skin response (GSR) is a physiological marker providing information about emotions, cognitive processes, and behavior [70]. GSR can potentially be used as a method of neuro-rehabilitation for people with mental disorders [71]. It can be used as a stress sensor to detect emotional states with a high success rate [72]. Adding GSR to eye tracking studies can help understand what moments of media content are more emotionally engaging for participants. It can also be used as a measure of workload or stress level. GSR, along with pupil dilation, reflects mechanisms of sympathetic activity, providing insights into a participant’s arousal level or cognitive load during a certain task. Overall, GSR is a useful tool that can be used to validate eye tracking and provide ground truth for cognitive and emotional processing. Therefore, the PredictEYE model, based on eye tracking data, used GSR data for its validation.

Feature engineering in eye tracking involves selecting and transforming relevant eye movement data attributes to improve the accuracy and effectiveness of the model in predicting mental states or other relevant outcomes. Many features can be extracted based on detecting events such as fixations, saccades, blinks, and features related to pupil size, gaze position, and areas of interest (AOIs). Effective eye-tracking feature engineering is essential for obtaining accurate and meaningful insights into mental states and other cognitive processes.

B. MODEL SELECTION

A detailed survey is performed to understand how statistical methods [73], [74], machine learning [20], [21], [22], [23], [24], [25], [75], deep learning [16], [20], [26], [27], [28], [29], and time series models [31], [76], [77], [78], [79] are applied to physiological signals to determine a person’s mental health.

The use of statistical methods is prevalent in mental health monitoring research. Techniques such as correlation analysis, linear regression models, t-tests, and analysis of variance (ANOVA) are commonly used to find associations and differences in the data. However, traditional statistical methods assume a linear relationship between variables and can be sensitive to outliers and non-normal data. Researchers have introduced robust statistical methods such as robust regression, trimmed means, and bootstrapping [73] to overcome these limitations. These methods can provide more accurate and reliable results, improving the validity and reliability of the analyses. Multiple regression and correlation analysis are powerful tools that help researchers analyze complex relationships between multiple variables [74]. In contrast, machine learning models such as neural networks and decision trees can handle non-linear relationships and outliers more effectively.

Statistical methods are also used in analyzing time series data. Traditional time series models can be linear or nonlinear, and some commonly used linear models include autoregressive (AR), moving average (MA), autoregressive moving average (ARMA), and autoregressive integrated moving average (ARIMA) [76]. Autocorrelation function (ACF) and partial autocorrelation function (PACF) analysis can help determine the appropriate model for a given time series dataset by showing how data sequences are related. Nonlinear models such as autoregressive conditional heteroskedasticity (ARCH), generalized ARCH (GARCH), exponential GARCH (EGARCH), threshold autoregressive (TAR), and nonlinear autoregressive (NAR) models are also available for improved analysis and prediction.

The use of machine learning and statistical analysis applied to physiological data has shown potential in monitoring mental health and assessing human mental states [20]. Machine learning models can be trained on labeled eye-tracking data to classify the user’s mental state [21], [22]. However, more investigation is required to develop the ideal eye-tracking paradigms and machine-learning algorithms for correctly diagnosing people with Autism Spectrum Disorder (ASD) and other neurological and neuropsychiatric illnesses.

Machine learning based on eye-tracking data could classify individuals with ASD and typically developing (TD) individuals with an 81% pooled accuracy [23]. A study on young children indicated that with an accuracy of 85.1 %, fixation times at the lips and body could significantly distinguish ASD from TD [24]. Video-based eye-tracking method for measuring brain function using readily available webcams has the potential for early detection, diagnosis, and

remote/serial monitoring of neurological and neuropsychiatric disorders [25].

Machine learning models, including Decision Trees, Naive Bayes, Support Vector Machines, and Random Forests, can analyze vast amounts of data and detect patterns [13] that may not be visible to human observers, providing an efficient health monitoring system. The studies showed that machine learning algorithms could analyze the data from these sources to accurately classify and predict psychological conditions, with classification accuracy ranging from 66% to 90% depending on the dataset and features used.

Recent technological advancements have enabled deep learning algorithms to automatically detect mental states, using various physiological signals such as visual metrics, EEG, and eye-tracking [16], [20]. One study used CNN-LSTM algorithms to analyze visual metrics time-series data and accurately classified individuals' mental health metrics levels with high accuracy, highlighting the potential benefits of home-based mental health monitoring for patients after oncologic surgery [16]. The other study proposed the EYE-CNN-DLSTM algorithm for psychological testing based on eye movement tracking data, which uses a fusion strategy combining CNN and DLSTM to evaluate patients with mental disorders [26].

A recent study proposed a deep learning-based technique that combines EEG and eye-tracking signals to improve emotion recognition accuracy [27]. The study demonstrated the advantages of using physiological signals to reflect a person's emotional state and highlighted the importance of eye-tracking signals in improving emotion recognition. The proposed approach utilizes a fusion model combining the Gaussian mixed model with signal filters, feature extraction techniques, and normalization methods to achieve precise emotion classification.

These algorithms, such as LSTM, can be optimized through hyperparameter tuning to predict better trends and fluctuations [28]. Additionally, deep learning models can learn and understand the mapping between inputs and outputs of linear and nonlinear models, supporting multivariate forecasting [29].

While deep learning and machine learning models have shown promising results in monitoring mental states, they are often limited by the lack of personalization. This means the models are trained on a general population and may not account for individual differences in behavior, preferences, and emotional responses. Personalized time series models can address this limitation by incorporating individual-level data and tailoring the model to each person's unique characteristics. This approach can improve the accuracy and effectiveness of mental health monitoring, especially in cases where individual differences are crucial, such as in psychiatric disorders. Additionally, deep learning models often require large amounts of data to train effectively, which can be challenging in mental health monitoring, where data collection is often limited. Personalized time series models can work with smaller datasets by incorporating

prior knowledge and individual-level characteristics, leading to more efficient and cost-effective mental health monitoring [80].

Time series modeling is a popular machine learning and deep learning-based technique that relies on historical data to predict future values of a target variable. The accuracy of attentional state classifiers can be increased by utilizing time series analysis for eye tracking data, followed by classification using convolutional neural networks as suggested in [31]. This method can improve mental health monitoring and diagnosis by optimizing attentional state classifiers. Another study [77] compares deep learning models for time series forecasting of COVID-19 cases and highlights the importance of historical data points and geographical location. These findings can inform policy development during the pandemic.

Time series data analysis has shown the potential to improve mental health outcomes and enhance online learning. Acikmese and Alptekin [78] proposed a method for predicting stress levels using mobile sensors and LSTM networks, while Zheng et al. [79] developed an approach for estimating engagement in online learning using webcams. These methods can help individuals manage their stress levels and improve the effectiveness of online learning.

1) SUMMARY

Time series analysis can help identify patterns and changes in data over time, allowing for predictive analytics and forecasting. Time series models are designed to capture temporal dynamics and dependencies, providing more accurate predictions of data that fluctuates over time [31]. Time series models can also be personalized to the individual, leading to more accurate predictions [16], [17], [18].

Eye tracking technology provides insights into participants' eye movements while engaging in various activities, including mental health monitoring [30]. Eye tracking can measure cognitive load, concentration, focus, drowsiness, consciousness, and other mental states important in identifying and managing mental health issues. Eye tracking is becoming increasingly important in mental health research, providing valuable information about emotional responses and mechanisms underlying human behavior [50], [52], [58], [59], [61], [64], [65].

Eye tracking with time series and personalized models offers several advantages in predicting a person's mental state. Firstly, time series analysis allows for identifying patterns and changes in attentional states over time, providing insights into attention and decision-making-related cognitive processes. Secondly, personalized models consider each individual's unique eye movement patterns and mental states, leading to more accurate predictions of their mental state than generic models. Thirdly, time series models capture temporal dynamics, account for individual differences, and handle missing data more effectively, resulting in more accurate and interpretable mental state predictions. Overall, eye tracking with time series and personalized models can provide a

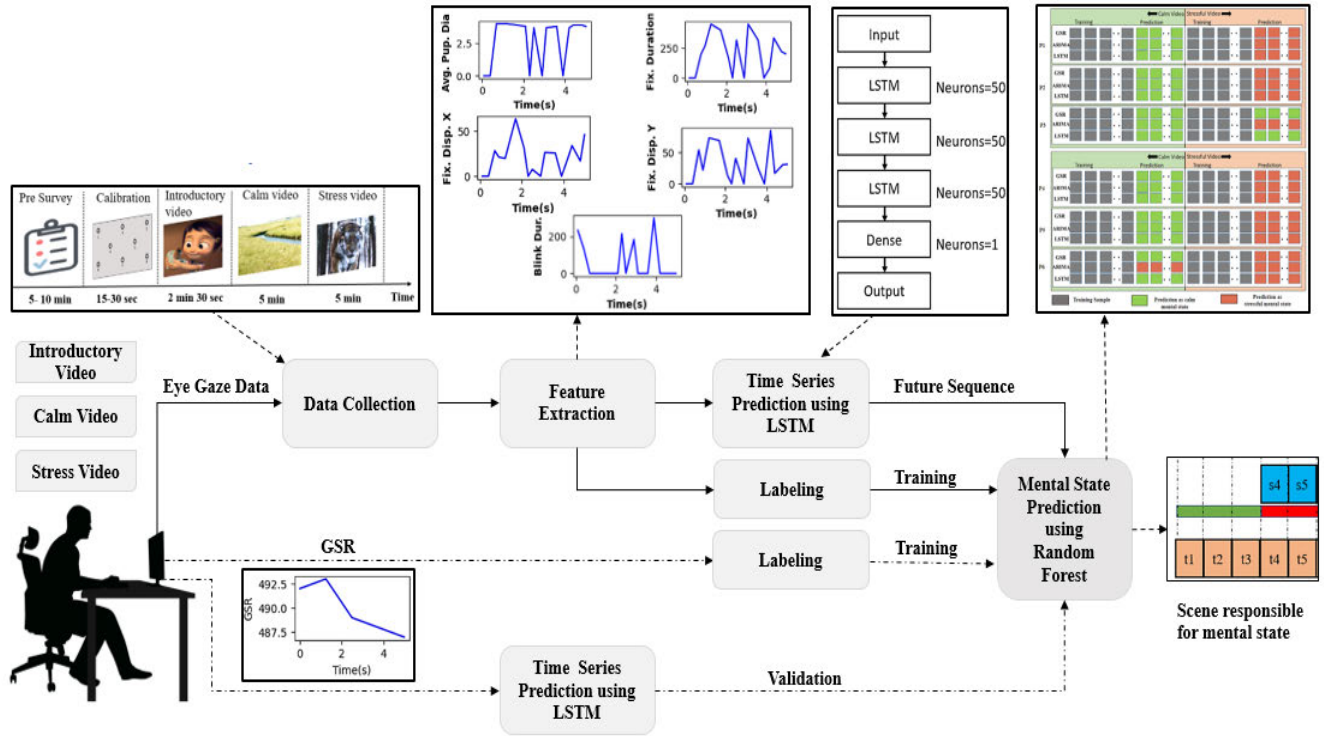


FIGURE 2. Architecture of PredictEYE, personalized time series model for mental state prediction.

powerful tool for developing accurate and personalized predictions of a person’s mental state.

Validation with GSR is important because it can provide insights into a participant’s cognitive and emotional processing, allowing for a more accurate interpretation of eye tracking data. The non-invasive nature of GSR measurement and its relative immunity to motion artifacts make it a valuable physiological measure for validating PredictEYE that predicts mental states [70], [71], [72].

Eye-tracking-based studies are a rapidly evolving field of research, and as such, there has been a limited amount of research on the use of time series models in this context.

III. SYSTEM MODEL

PredictEYE is a personalized time series model designed to forecast future sequences by capturing the trends present in the past eye tracking data and to predict the mental state of an individual and the scene responsible for the mental state [30], as illustrated in Figure 2. PredictEYE follows a pipeline that involves collecting and processing raw eye gaze data, extracting features, using an LSTM module to predict future feature sequences, training a Random Forest model on labeled data, and finally using the LSTM predictions as input to the Random Forest model to make predictions on mental state.

Similarly, GSR data is collected and processed using the same pipeline, including LSTM prediction and Random Forest model training. The predicted mental state from the Random Forest model based on GSR is used to validate the

predictions made by the PredictEYE system based on eye tracking measures.

A. DATA COLLECTION

A 10-minute video including 5 minutes of calm and 5 minutes of stressful video was utilized as a stimulus [81]. Before the actual stimulus, an introductory segment with a short animated movie was played for a 2-minute and 30 seconds. The calm [82] and stressful video scenes [83] were presented subsequently. The calm segment of the video was selected based on its ability to induce relaxation within 5 minutes, as per the recommendation of specialists in stress and anxiety therapy. The introductory video was not included in the analysis but served to acquaint the participants with the experimental setup. The calm video period was considered the baseline phase, with the expectation that the participants would experience a state of relaxation during this time. The data collection and feature extraction procedures are explained in Figure 3. The sampling frequency of 60 Hz allowed us to capture 60 samples of raw eye gaze data per second. In total, accumulated a substantial dataset of observations consisting of 10 minutes, or 600 seconds. This translates to a vast raw eye gaze data repository comprising 36,000 samples for a participant. The raw eye gaze data includes (X, Y) gaze coordinates, pupil diameter, and timestamp.

B. FEATURE EXTRACTION AND LABELING OF THE DATA

The features were extracted from raw eye gaze data using BeGaze 3.7 software, which employs the dispersion-based

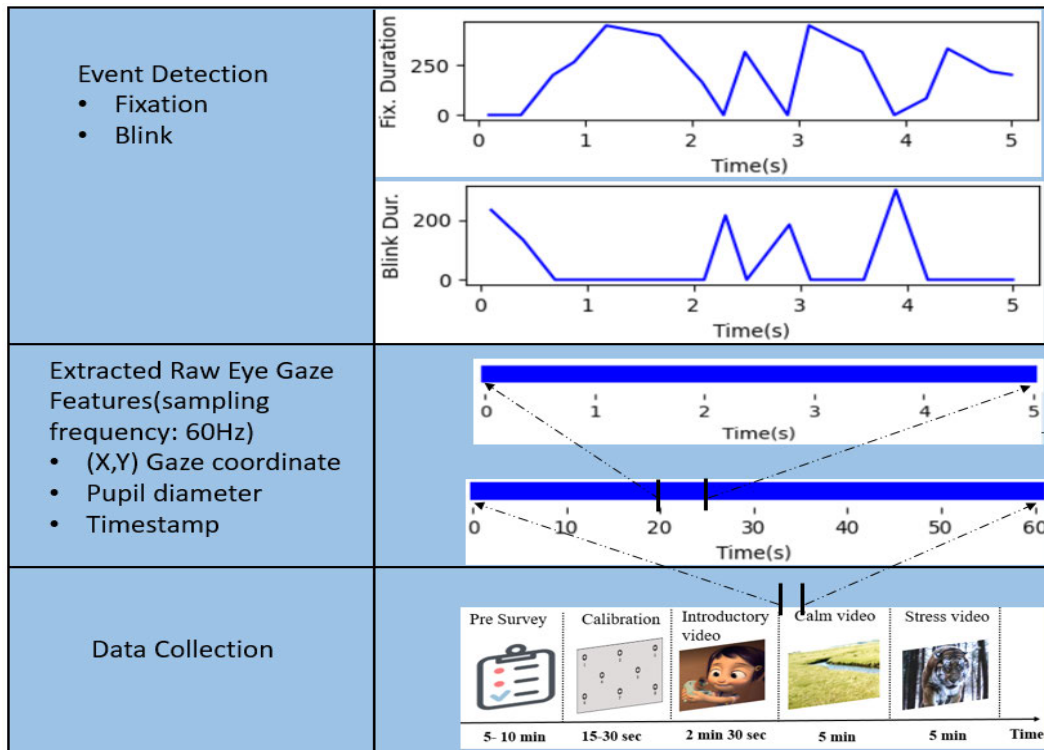


FIGURE 3. Feature extraction procedures.

I-DT algorithm for detecting events such as fixation and blink [84]. The identification by dispersion threshold (I-DT) algorithm is designed to detect fixation events in eye-tracking data using two critical parameters: the dispersion and duration threshold. It considers the X and Y coordinates from raw eye gaze data and identifies fixations as clusters of gaze points with a low dispersion value within a specific time interval. The algorithm can accurately identify and locate fixations throughout the eye-tracking data by iteratively applying this process. A minimum duration of 80 ms and a maximum dispersion of 100 pixels were considered in the I-DT algorithm for the fixation detection. A minimum duration of 70 ms was considered for the detection of blink. Events like fixation and blink have been identified, and the corresponding features have been extracted. The features extracted were Fixation duration, Fixation Dispersion X-Axis and Y-Axis, Pupil Diameter, and Blink Duration. In eye movement data analysis, fixations and blinks are mutually exclusive events. When fixation-related features are considered at a given time 't', blink-related features have a value of zero, and vice versa, due to the closure of eyes during blinks and the inability to have fixations during those periods.

Fixation Duration: Fixation duration is the amount of time that a person's eyes remain still while they are fixating on a particular object. This measurement is typically recorded in milliseconds (ms) and determines how long a person is paying attention to a particular object or area of interest.

Fixation Dispersion X-Axis and Y-Axis: Fixation dispersion measures how widely dispersed a person's eye

movements are while they are fixating on a particular object or scene. It is typically measured in terms of the X-axis (horizontal) and the Y-axis (vertical) in pixels. Fixation dispersion in the X and Y axes provides insights into how participants distribute their attention within an image or scene during fixation. It quantifies the spatial variability in gaze patterns, helping researchers understand how individuals allocate their focus during different mental states, whether calm or stressful, within visual stimuli.

Pupil Diameter: Pupil diameter measures the size of a person's pupils and is often used to indicate cognitive load or arousal. The size of the pupils can change in response to changes in visual stimuli and cognitive processing demands, and it is typically measured in millimeters (mm).

Blink Duration: Blink duration is the amount of time a person's eyes are closed during a blink. This measurement is typically recorded in milliseconds and can provide information about a person's cognitive workload, fatigue, or attentional focus.

Fixation, blink-based features, and pupil diameter were input to the LSTM model. The eye tracking data obtained from each participant during calm and stress videos are labeled and provided as input to the LSTM, followed by a Random Forest algorithm to predict mental state.

The plot of features extracted from eye tracking data for Participant 1 while watching the stress video is shown in Figure 2 as part of feature extraction. Each feature in the Y axis is plotted across the starting time of that particular event.

The data is labeled based on a threshold approach during the calm video period, which is considered the baseline phase. Once the threshold value is determined, any value above the threshold is labeled as 'stressful', and any value below the threshold is labeled as 'calm'. This approach assumes that the baseline phase represents a state of relaxation, and any deviation from this state can be considered an indicator of stress.

C. TIME SERIES PREDICTION USING LSTM

The input data for the task at hand comprises time series eye tracking features. These features were preprocessed and subsequently input into a univariate LSTM (Long Short-Term Memory) model. A popular time series forecasting method is LSTM, which is based on recurrent neural networks (RNN). LSTM uses memory cells to remember the previous stages in the network [85]. The memory cells contain the input, forget, and output gates. The input gate controls the flow of input activation, the forget gate decides how long the value should remain in the cell, and the output gate controls the flow of cell activation into other networks. The LSTM model is designed to learn from the patterns and dependencies present in the input data. By processing the time series input data, the model learns to capture the temporal dynamics present in the data and uses this information to make predictions about future values of the features.

The performance of time series models can be evaluated by comparing the predicted values with the actual values of the target variable. Commonly used metrics for evaluating time series models include:

- 1) Mean Absolute Error (MAE)- This is the average of the absolute differences between the predicted values and the actual values. Since it finds the absolute value, the value of MAE will be positive. A lower value indicates better accuracy, as it means the model is closer to the true values. MAE is an easily interpretable measure of accuracy that is widely used in time series forecasting.
- 2) Residual sum of squares (RSS) - It is the sum of squares of the residuals. A lower value indicates the model is closer to the true values and will improve the accuracy. RSS is a measure of the overall fit of the model and is often used to compare the performance of different models.
- 3) Mean Squared Error (MSE)- The average of squared errors. Since it squares the errors, the value of MSE will be positive. A lower value indicates better accuracy. MSE is a widely used measure of accuracy but less interpretable than MAE.
- 4) Root Mean Squared Error (RMSE)- It is the square root of MSE. A lower value indicates better accuracy, as it means the model is closer to the true values. RMSE is a popular measure of accuracy because it is interpretable and easy to compare across models.
- 5) Mean Absolute Percentage Error (MAPE)-It is a metric used to evaluate the accuracy of a predictive model. It calculates the average percentage difference between

the predicted and actual values, providing a relative measure of accuracy that is easy to interpret. In time series forecasting, MAPE is a popular metric because it provides a simple way to assess the model's accuracy. A lower value of MAPE indicates better accuracy, meaning that the model is closer to the actual values.

- 6) Mean Error (ME) represents the average difference between actual and predicted values. A value of zero indicates that the model is unbiased. ME is less commonly used than other measures of accuracy, but it can provide useful information about the direction of errors.
- 7) Mean Percentage Error (MPE)- The average percentage of actual values deviating from predictions. A value of zero indicates that the model is unbiased. It can provide useful information about the direction and magnitude of errors.

These error statistics are calculated based on the formula (1) to (7).

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (1)$$

$$RSS = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (5)$$

$$ME = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i) \quad (6)$$

$$MPE = \frac{1}{n} \sum_{i=1}^n \left(\frac{y_i - \hat{y}_i}{y_i} \right) \times 100\% \quad (7)$$

In the given formulas, y_i represents the actual values, and \hat{y}_i represents the predicted values.

By comparing the models based on these metrics, researchers can gain insight into the models that perform better for a specific task.

D. MENTAL STATE PREDICTION WITH RANDOM FOREST ALGORITHM

The Random Forest algorithm is trained based on the labeled eye tracking data. The predicted sequences from the LSTM model were given as input to the Random Forest algorithm to classify the mental state based on the predicted future data sequence. Being an ensemble learning method, Random Forest combines multiple decision trees to generate a prediction. In our study, the Random Forest algorithm was used as the classification algorithm to classify the mental

state. Each decision tree created by the Random Forest algorithm is trained using a different subset of the data. The forecasts of each decision tree are then combined to get the final prediction.

By analyzing eye-tracking data, PredictEYE can also identify the areas of the video that an individual was looking at during different stages of the video and correlate these areas with the individual's reported mental state. This information can provide valuable insights into factors contributing to an individual's mental state, allowing for more targeted interventions or treatments.

Classification model performance is assessed according to various criteria, including accuracy, precision, recall, F1 score, and area under the receiver operating characteristic (ROC) curve. These metrics judge how well the model can place instances in the appropriate classes. Accuracy is the ratio of correctly predicted instances to the total instances. The precision metric measures the proportion of true positives (i.e., correctly detected cases) among all instances classified as positive. In contrast, the fraction of true positives that were correctly identified out of all positive instances is measured by the recall. F1 score is a combined metric that considers both precision and recall to provide an overall measure of the model's performance. The ROC curve is a graphical representation of the trade-off between a classifier's true positive rate (TPR) and the false positive rate (FPR). A high Area Under the Curve (AUC) indicates that the classifier is able to discriminate well between positive and negative examples, making it a useful metric for many applications.

The accuracy, precision, recall, F1 Score, TPR, and FPR of the PredictEYE model are calculated based on the formula (8) to (13), respectively.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TN} + \text{TP} + \text{FN} + \text{FP}} \times 100\% \quad (8)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\% \quad (9)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\% \quad (10)$$

$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\% \quad (11)$$

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (12)$$

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \quad (13)$$

where TP is True Positive, TN is True Negative, FP is False positive, and FN is False Negatives.

The importance of accuracy, precision, recall, F1 score, and ROC lies in their ability to evaluate the performance of a classification model based on different aspects. Accuracy provides an overall measure of the model's performance, while precision and recall help identify the model's ability

to correctly identify instances of a specific class. The F1 score combines both precision and recall to provide a comprehensive evaluation of the model's performance.

E. VALIDATION WITH GSR

Galvanic Skin Response (GSR) is a non-invasive physiological measure that reflects the activity of the sympathetic nervous system and has been widely used in research related to emotional and cognitive processing [71]. GSR is easy to measure and useful for studying mental states in various settings. Compared to other physiological measures, such as heart rate or electroencephalography (EEG), GSR is less affected by motion artifacts and has a slower response time [70]. GSR can be useful for validating models that predict mental states [86].

The performance of PredictEYE was validated by collecting GSR data from participants as they watched calm and stressful videos while their eye movements were tracked. An LSTM model was trained on this data to predict future sequences of GSR data. In parallel, a Random Forest model was trained to predict the participants' mental state based on the predicted GSR sequence. The Random Forest predictions based on the GSR data were used to validate the mental state predictions made using eye tracking data.

IV. IMPLEMENTATION

The experimental study was performed on hospital employees ($n=6$, 3 male, Mean age = 33.5, SD=5.6, age range = 26 to 42) [30]. The procedures followed in the experimental research are shown in Figure 2. The participant was informed about the data collection processes and obtained written consent. The eye tracking data was collected using the SMI Redn Professional Eye Tracker (Company: SensoMotoric Instruments, Germany) with a sampling frequency of 60Hz and Experiment Center 3.7 software, which provides comprehensive tools for stimulus presentation and precise data collection. The eye tracker was connected to the laptop, and the participant was asked to sit comfortably in front of the system at an appropriate distance. The distance between the participant's eye and the eye tracker was maintained within the limit of 50cm. The eye tracking experiment was conducted in a controlled environment, with a standardized lighting setup consistently maintained across all participants and video stimuli. The display screen brightness, color, and contrast were predefined and uniformly maintained for all participants throughout the study. The eye tracker was calibrated before starting each experiment. The data was collected for a duration of 10 minutes for each participant. Over the 10-minute data collection period for all participants, the eye tracker recorded 216,000 eye movement data samples.

GSR data was collected using a grove-GSR sensor, with two electrodes attached to two fingers of a hand, [87]. It measures the electrical resistance of the subject's skin, and this information is then used to produce an output voltage, typically measured in millivolts (mV). The participant

was given the wearable band and instructed to wear it in accordance with the guidelines. Arduino integrated development environment (IDE) software was used to collect GSR data. Eye tracking and GSR data were collected using the same machine, ensuring the system clock remained consistent. To synchronize the data collection, a software trigger was set to initiate eye tracking and GSR measurements simultaneously, resulting in timestamps that are accurately aligned across the datasets.

The collected eye tracking time series features were fed into the LSTM model to predict the new time series data. A sequence-to-sequence regression LSTM network was used to predict future time series values. The training data was initially normalized to have a mean of zero and a variance of one. The model's predictors and responses were then ready for training. The training sequence with values altered by a single step was used as the response. At each stage of the input sequence, the LSTM network gains the ability to predict the value of the subsequent step. The sequence of data without the final time step is called a predictor.

The LSTM model, as shown in Figure 2, has 3 LSTM layers followed by a Dense output layer. Each LSTM layer has 50 hidden units or neurons that help to store and manipulate information over time. The output of the third LSTM layer is then fed to the Dense, fully connected layer, which has a single output unit. This output unit predicts the next value in the time series based on the input sequence of window-size time steps.

It takes a sequence of window-size time steps and each extracted eye gaze feature as input. The window size indicates the number of consecutive data points considered at a time when performing calculations or making predictions. The window size is a hyperparameter that defines the number of time steps (or data points) to include in each input sample to the LSTM model. The SMI eye tracker used for the data collection had a sampling frequency of 60 Hz. Initially, a window size of 60 was considered to align with the data frequency, but it caused a loss of detail during rapid eye movements. The window size was reduced to 10 to preserve more information, representing 1/6th of a second interval. This allowed for capturing finer changes in gaze behavior and improved the model's prediction accuracy for future eye movements and reactions [88]. A window size that is too small (less than 10) in time series analysis can lead to significant issues, including loss of information, increased noise, and overfitting. A value of 10 demonstrated superior performance through experimentation with different window sizes, effectively balancing information capture and overfitting prevention in the time series analysis. The first input sample to the LSTM model will contain the feature values from time step 1 to 10, and the second input sample will contain the values from time step 2 to 11, and so on, as shown in Figure 4. The model is trained to predict the subsequent value based on the previous ten values in each sequence. The sliding window approach is used to create these input-output pairs.

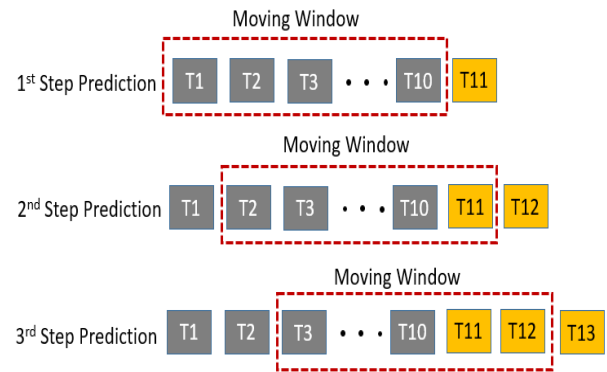


FIGURE 4. Time series prediction.

The activation function used in the LSTM layer was ReLU (rectified linear unit). ReLU is a popular activation function in deep learning models that has been proven effective in real-world applications. The network parameters were optimized using the Adam optimizer [89], known for its quick and efficient convergence.

The loss function used for training the model was mean squared error (MSE). The network was trained for 100 epochs, with a batch size of 16, and a gradient threshold of 1 to avoid gradient explosions. The initial learning rate was set to 0.005, and a factor dropped to 0.2 after every 125 epochs. This allowed the network to converge smoothly and avoid getting stuck in local minima.

During training, the model was fed with input sequences of window-size time steps and one feature and was trained to predict the next value in the time series. After training, the model was evaluated on a test set using the same window size, and the MSE loss was computed. Finally, the model generated a future sequence by feeding in the last window-size time steps of the training data and iteratively predicted the next value in the time series. The LSTM network was trained to capture the dependencies and patterns in the input time series data, which can be used to make accurate predictions about future eye movements.

The Random Forest algorithm consists of several key components. First, the algorithm randomly selects a subset of features from the input data at each split point. This aids in lowering overfitting and enhancing the model's precision. Second, the method constructs several decision trees from various subsets of the data. This enhances the model's overall accuracy by capturing various data features. The algorithm then integrates all the different decision tree predictions to arrive at a final prediction.

The Random Forest algorithm is trained on labeled data, and once the training is complete, it is applied to the predicted time series data to predict each participant's mental state. Since the physiological data is unique to each person, the Random Forest algorithm is trained separately for each individual. The algorithm uses the predicted data sequence to classify the person's mental state as calm or stressed.

The total video-watching time was divided into ten groups. After each group, the future sequence was predicted with

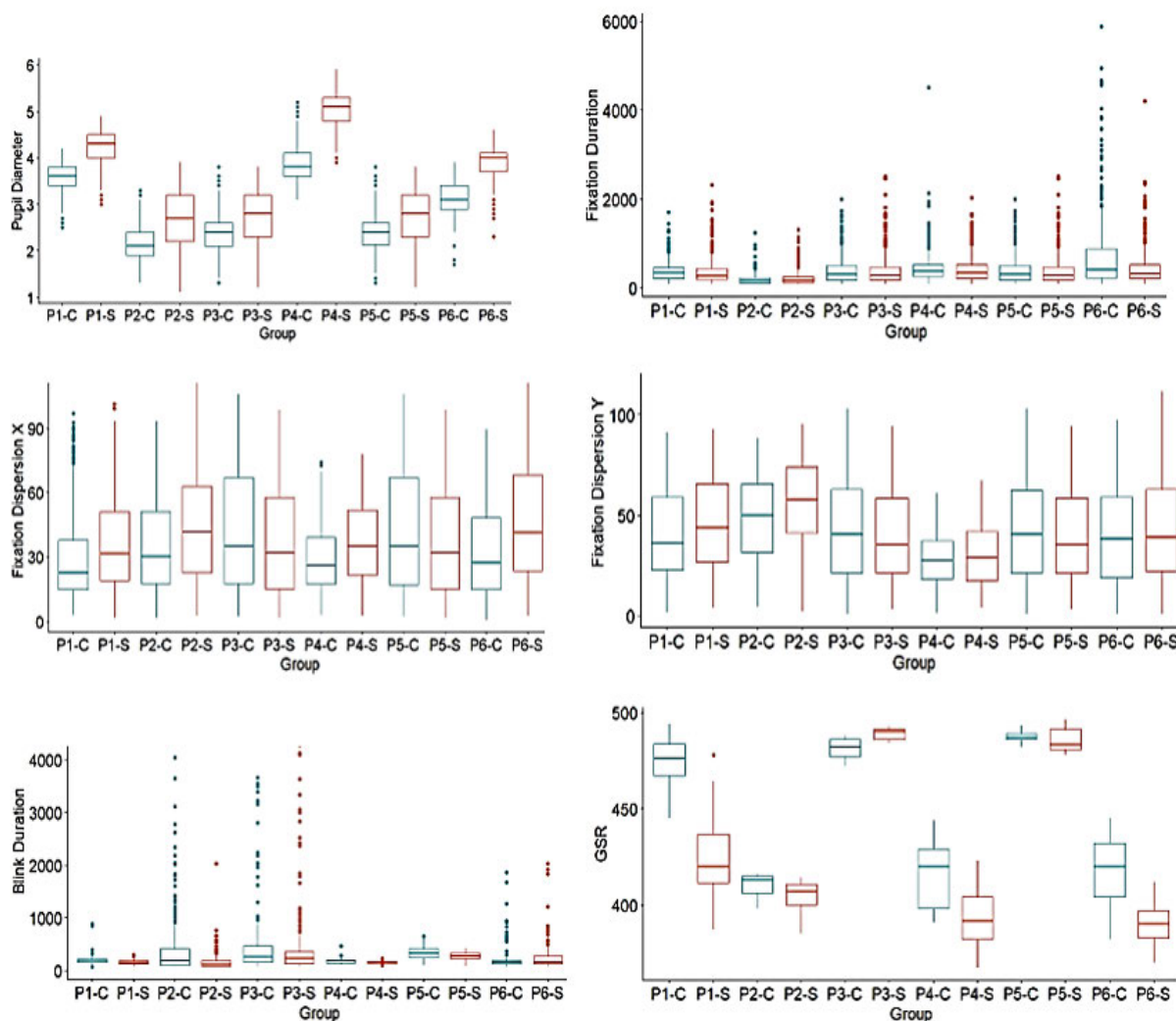


FIGURE 5. Boxplot of the eye tracking measures.

LSTM. Thus, the mental state prediction was performed every minute to forecast the expected sequence of mental states that might be observed in the subsequent video-watching period. This approach allowed for identifying patterns and trends in mental state changes over time, which could inform the development of more effective interventions or personalized treatments for individuals with mental health concerns.

The code implements LSTM architecture using Keras in Python and imports various libraries, including numPy, pandas, matplotlib, and scikit-learn. The input data is normalized using the MinMaxScaler function. Additionally, a Random Forest algorithm is also implemented using the Scikit-learn library.

V. RESULT ANALYSIS

The PredictEYE is a personalized time series regression model designed to forecast an individual’s mental state based

on eye movements. It also finds the scene responsible for that mental state. PredictEYE has been analyzed in two stages: data exploration and statistical analysis, and performance evaluation of the PredictEYE model in estimating the mental state.

Data exploration and statistical analysis were the first stages of analysis of the PredictEYE model. This stage involved analyzing the collected data to understand the distribution and variability of the dataset. This analysis helped to identify patterns, trends, and relationships between different variables in the dataset, which is crucial for developing accurate and reliable predictive models. After the initial data exploration and statistical analysis, the next step is to evaluate the PredictEYE model’s performance in forecasting the future sequence. This is important to measure the model’s accuracy in predicting an individual’s eye gaze sequence. The third stage of analysis involves the performance evaluation of the classification model in estimating the mental state. In this stage, the model is trained

TABLE 1. Results of welch two sample t-test.

Participant	Feature Name	P-Value	Calm(mean)	Calm(SD)	Stressful(mean)	Stressful(SD)	Segment length
P1	Avg Pupil Diameter (mm)	<0.0001	2.36	1.73	3.23	1.79	169
	Fixation Duration (ms)	0.91	204.44	199.80	202.00	163.91	169
	Fixation Dispersion X (pixels)	0.0001	16.79	19.42	26.55	23.68	169
	Fixation Dispersion Y (pixels)	0.75	33.09	31.34	34.19	27.75	169
	Blink Duration (ms)	0.0009	67.96	97.60	35.33	68.16	169
	GSR (mV)	<0.0001	484.07	7.23	419.66	13.62	60
P2	Avg Pupil Diameter (mm)	<0.0001	0.65	0.93	1.38	1.42	166
	Fixation Duration (ms)	0.15	66.98	115.33	86.05	114.97	166
	Fixation Dispersion X (pixels)	<0.0001	12.24	21.54	25.85	30.94	166
	Fixation Dispersion Y (pixels)	0.01	19.72	30.19	29.25	32.49	166
	Blink Duration (ms)	<0.0001	297.70	544.14	90.45	193.49	166
	GSR (mV)	<0.0001	406.19	4.36	399.33	8.20	60
P3	Avg Pupil Diameter (mm)	0.03	1.69	1.17	1.32	1.38	111
	Fixation Duration (ms)	0.60	289.57	320.19	265.15	397.83	111
	Fixation Dispersion X (pixels)	0.50	27.42	28.72	24.84	30.37	111
	Fixation Dispersion Y (pixels)	0.04	29.54	28.58	22.03	26.91	111
	Blink Duration (ms)	0.70	293.43	1535.25	356.22	805.79	111
	GSR (mV)	0.25	486.45	1.394538	485.35	3.950683	60
P4	Avg Pupil Diameter (mm)	<0.0001	3.61	1.09	4.67	1.25	140
	Fixation Duration (ms)	0.44	336.99	220.02	314.94	256.04	140
	Fixation Dispersion X (pixels)	<0.0001	26.56	14.78	35.35	18.96	140
	Fixation Dispersion Y (pixels)	0.03	29.54	28.58	21.42	26.58	140
	Blink Duration (ms)	0.41	15.10	52.24	10.51	40.26	140
	GSR (mV)	<0.0001	407.71	14.08	389.49	12.62	60
P5	Avg Pupil Diameter (mm)	<0.0001	4.09	0.85	4.59	0.79	135
	Fixation Duration (ms)	<0.0001	519.74	532.02	310.33	262.33	135
	Fixation Dispersion X (pixels)	0.002	26.43	15.10	32.79	17.85	135
	Fixation Dispersion Y (pixels)	0.22	18.39	11.03	16.83	10.30	135
	Blink Duration (ms)	0.49	13.19	77.76	7.77	47.35	135
	GSR (mV)	<0.0001	489.98	2.06	483.40	3.55	60
P6	Avg Pupil Diameter (mm)	<0.0001	2.20	1.62	3.60	1.16	150
	Fixation Duration (ms)	0.31	388.23	542.96	324.16	279.60	150
	Fixation Dispersion X (pixels)	<0.0001	20.52	22.81	48.12	31.40	150
	Fixation Dispersion Y (pixels)	0.04	27.05	27.30	35.14	27.66	150
	Blink Duration (ms)	0.03	124.75	314.60	40.55	190.61	150
	GSR (mV)	<0.0001	403	14.01	386.47	7.12	60

to classify the mental state of an individual based on their eye gaze sequence.

A. DATA EXPLORATION AND STATISTICAL ANALYSIS

The analysis of the mean of all eye tracking measurements taken while watching the relaxing and stressful videos by all participants is shown in Figure 5. 'P1-C' represents Participant 1's data during the calm video-watching duration, and 'P1-S' represents Participant 1's data during the stressful video-watching duration. The study found that the mean of pupil diameter and fixation dispersion in the X and Y axis increased during stress video watching compared to calm video watching, while the mean of fixation duration and blink duration decreased. These observations suggest that eye-tracking features are potentially useful for detecting changes in mental state, particularly during stress-inducing

tasks. Additionally, it was noted that sweat glands become active during periods of stress or emotional upset and secrete more moisture into the body, resulting in increased conductivity and decreased resistivity. As resistivity is measured using sensors, a decreased GSR value would be expected if the individual is stressed. This pattern was observed in most participants when analyzing their GSR values. GSR offers a more direct measurement of arousal, reflecting physiological responses, while eye tracking data indirectly measures visual attention and gaze behavior. Even subtle changes in eye tracking data can be valuable, as they can reveal how individuals respond emotionally to different stimuli.

Participant 3 exhibited a significant difference in the measured variables compared to the other participants. Notably, no significant difference was observed in this participant's body resistivity between stress and calm video-watching

sessions, suggesting that they did not experience a significant increase in stress while watching the stress video, according to the ground truth. Eye tracking measures supported this finding, with a variation in certain features observed for this participant compared to others.

To ascertain whether there was a significant difference between participants in their eye tracking data while watching calm and stressful videos, a Welch two-sample t-test was performed on all features. It is a statistical test that is used to compare the means of two independent groups when the variances of the two groups are not equal. We aimed to investigate the mental state of individuals at the end of watching calm and stressful videos. Each video lasted for 5 minutes, but for analysis, we focused on the last minute of each video. This statistical test evaluates potential differences in eye movement patterns between the two types of videos. Analyzing the data allows for the exploration of potential variations in ocular behaviors in response to different video content, providing valuable insights into how these videos influence the mental state of participants during the observation.

The null hypothesis was set as there is no significant difference between the two types of data, while the alternative hypothesis stated that there was a considerable difference between the calm and stress data. This test was performed to statistically evaluate whether the observed differences in the features between the two types of video data were likely to have occurred by chance or if they were indicative of a significant difference between the two states. Table 1 shows the P-value obtained on performing the Welch Two-sample t-test on all the features. Based on the results of the Welch Two Sample t-test, it can be concluded that there is a significant difference between the calm and stressful video data for most of the features, except for Participant 3. The null hypothesis can be rejected, and the alternative hypothesis is supported, indicating that there was not much substantial difference after watching the stressful video compared to the calm video for most of the analyzed features of Participant 3.

The fact that the p-values for most participants were less than 0.05 for most of the features suggests that the differences between the calm and stressful video data were significant and not simply due to chance. However, for Participant 3, the P-value obtained for the features fixation duration, fixation dispersion X-axis, and blink duration at the end of the video-watching time of calm and stressful videos were not less than 0.05. This suggests that no significant difference was observed in these features at the end of this participant's calm and stressful video-watching sessions. Additionally, the ground truth GSR values for this participant, at the end of the video-watching time of calm and stressful videos, also had a P-value of not less than 0.05, indicating that there was no significant difference in GSR between the calm and stressful video-watching sessions.

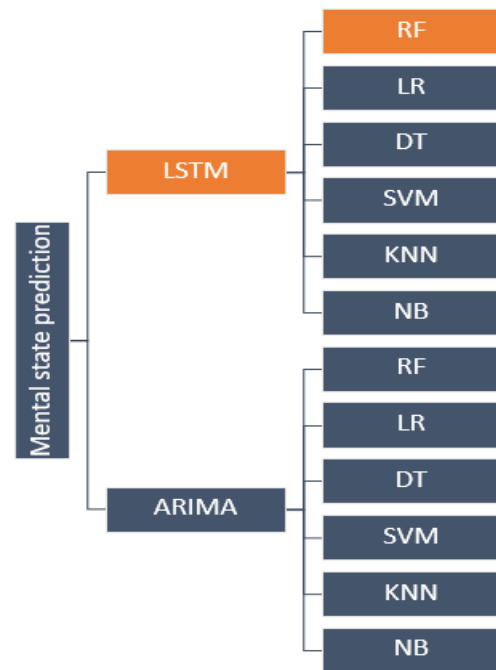


FIGURE 6. Comparison of PredictEYE with other models.

B. PERFORMANCE EVALUATION OF PREDICTEYE

During the data collection phase, a total of 216,000 data samples were gathered from all the participants over a span of 10 minutes. Each participant contributed an average of 36,000 data samples, as they were observed while watching both calm and stressful videos. These data samples were then utilized to make predictions regarding the participants' responses, which were subsequently analyzed for further insights. The correctness of the predicted data sequence was evaluated with standard performance measures such as Mean Error, Mean Absolute Error, Mean Percentage Error, Mean Absolute Percentage Error, Mean Squared Error, and Root Mean Squared Error. The prediction accuracy and the error statistics are analyzed based on various performance measures.

The PredictEYE model, a combination of LSTM and Random Forest models, was compared with various combinations of models, as depicted in Figure 6. The evaluation process began by comparing the time series predictions of LSTM with ARIMA predictions. Subsequently, the mental state predictions based on Random Forest were assessed across multiple machine learning algorithms, including Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), K Nearest Neighbor (KNN), and Naive Bayes (NB). The following section presents a detailed comparison of the PredictEYE model with different models, highlighting its performance and effectiveness in predicting mental states.

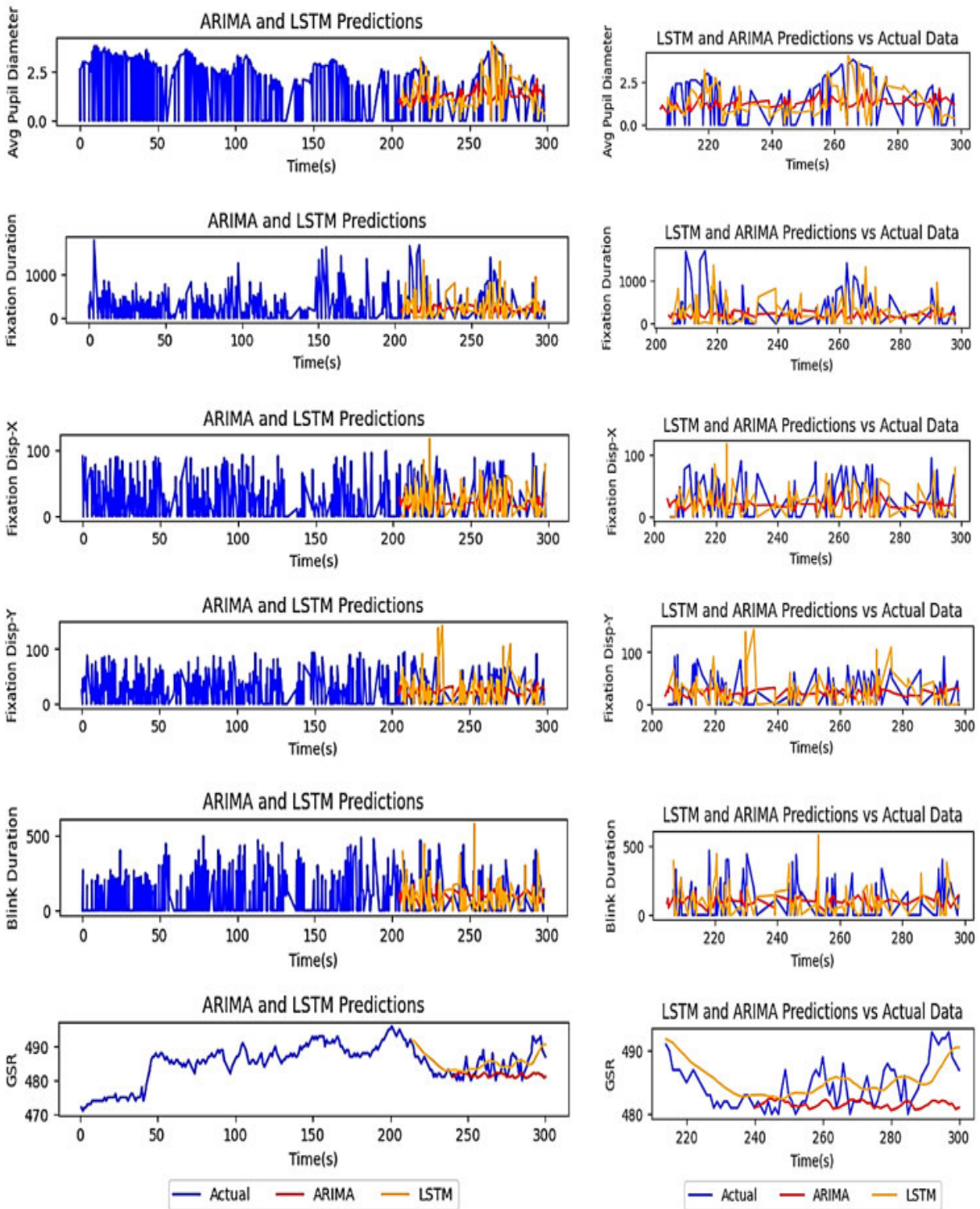


FIGURE 7. Forecasting data sequence of participant-3 after watching stressful video using LSTM and ARIMA models in PredictEYE. Fixation Disp-X is fixation dispersion X and fixation Disp-Y is fixation dispersion Y.

1) PERFORMANCE EVALUATION OF LSTM WITH ARIMA BASED ON ERROR STATISTICS

The ARIMA model is a time series analysis algorithm that aims to understand trends, seasonality, and cyclic behavior in

the data [90]. Exploratory data analysis is performed before fitting the model, which involves checking for autocorrelation between current and past values at different lags. This information is used to select the appropriate ARIMA model

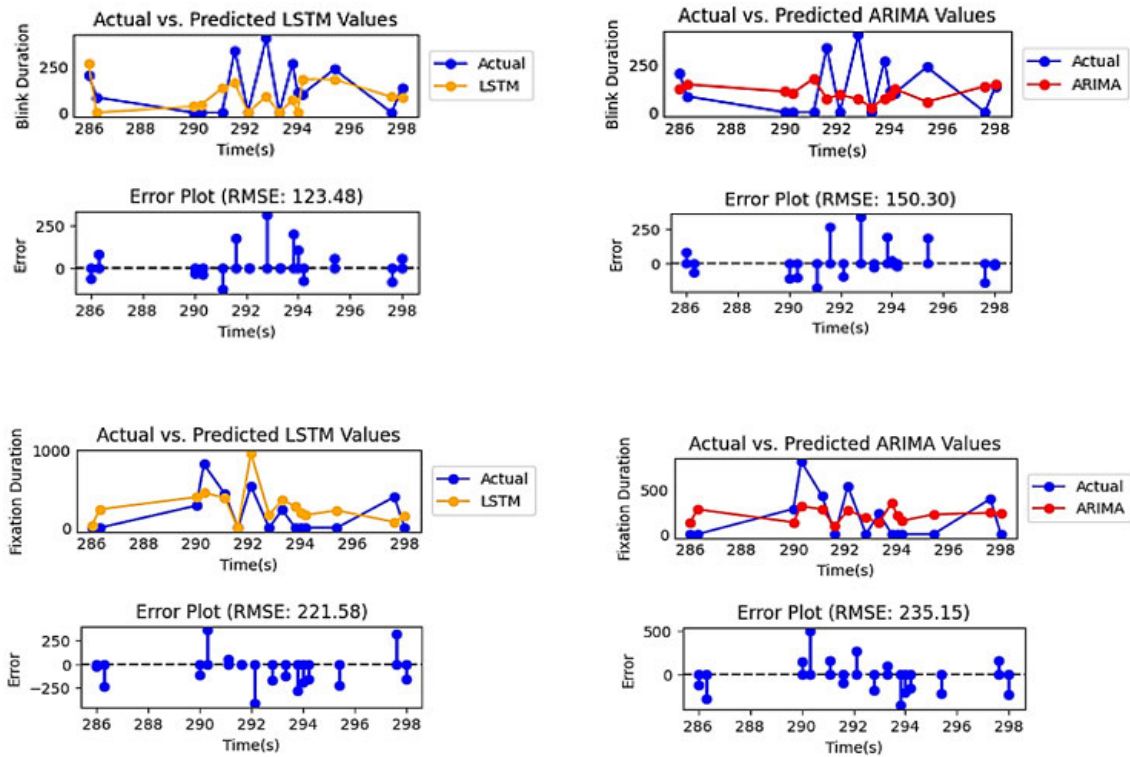


FIGURE 8. Comparison between actual and predicted values of LSTM and ARIMA for a short interval.

TABLE 2. Error statistics based on the prediction while watching calm video.

Model	Features	Performance measures						
		MAE	RSS	MSE	RMSE	MAPE	ME	MPE
LSTM	Pupil Diameter (mm)	0.34	10.55	0.19	0.42	0.19	0.05	-0.01
	Fixation Duration (ms)	139.66	1689383.93	29638.31	172.16	71.77	29.41	51.25
	Fixation Dispersion X (pixels)	14.86	17837.80	312.94	17.69	133.27	8.38	121.79
	Fixation Dispersion Y (pixels)	21.2	33705.17	591.31	24.31	60.84	-5.79	23.72
	Blink Duration (ms)	41.81	131132.49	2731.92	52.26	34.46	22.09	26.05
	GSR (mV)	1.42	21.17	4.23	2.05	0.30	0.98	0.21
ARIMA	Pupil Diameter (mm)	0.73	38.82	0.57	0.76	19.84	-0.73	-19.84
	Fixation Duration (ms)	171.86	4406649	73884.91	255.19	1.05	-51.89	0.22
	Fixation Dispersion X (pixels)	15.33	23562.35	351.67	18.75	123.89	4.87	105.62
	Fixation Dispersion Y (pixels)	21.76	41707.75	622.50	24.95	56.42	-9.16	12.54
	Blink Duration (ms)	50.25	136372.19	3588.74	59.90	42.26	36.42	36.74
	GSR (mV)	5.2	386.18	43.31	5.76	0.14	-4.9	-0.01

for the data, and the ACF and PACF help to understand the model’s autoregressive, moving average, and integrated components. The best-fit ARIMA model can then be used to forecast the next successive values of eye tracking features.

The predicted sequences of LSTM and ARIMA models at different intervals while watching calm and stressful videos were compared with the actual values to understand prediction accuracy. After watching stressful videos, the predicted sequence of LSTM and ARIMA models are shown in Figure 7. The plot illustrates the results based on 80% of the data used for training while watching the stressful

video, while the remaining 20% of the data was used for prediction with LSTM and ARIMA methods. The figure displays the performance and predictive capabilities of these models on the given dataset. Figure 8 compares the actual and predicted values of features ‘blink duration’ and ‘fixation duration’ based on LSTM and ARIMA predictions for a short time interval. The plot provides clear evidence that fixations occur exclusively during the absence of blinks. The reduction in RMSE in the LSTM model, as demonstrated in Figure 8, carries substantial implications for model performance assessment. A lower RMSE indicates that the

TABLE 3. Error statistics based on the prediction while watching stress video.

Model	Features	Performance measures						
		MAE	RSS	MSE	RMSE	MAPE	ME	MPE
LSTM	Pupil Diameter (mm)	0.22	4.44	0.06	0.25	5.56	0.08	2.36
	Fixation Duration (ms)	102	883989.32	15785.52	125.64	54.21	24.21	36.6
	Fixation Dispersion X (pixels)	16.49	23618.86	421.76	20.53	96.23	-0.30	67.91
	Fixation Dispersion Y (pixels)	16.79	24897.53	440.61	20.25	1.27	-0.78	0.37
	Blink Duration (ms)	41.62	28188.9	2818.89	53.09	23.42	-2.97	6.14
	GSR (mV)	3.83	250.43	28.92	4.52	0.10	-1.61	-0.01
ARIMA	Pupil Diameter (mm)	0.38	13.69	0.31	0.45	0.24	-0.05	-0.01
	Fixation Duration (ms)	171.86	4406649	73884.91	255.19	1.05	-51.89	0.22
	Fixation Dispersion X (pixels)	17.38	29049.29	440.14	20.97	109.35	5.75	91.88
	Fixation Dispersion Y (pixels)	17.83	30656.09	559.01	22.28	0.73	-1.49	-0.23
	Blink Duration (ms)	45.66	65777.52	3288.87	57.34	35.83	8.3	20.3
	GSR (mV)	9.26	1168.63	106.23	10.3	2.09	-9.26	-2.09

LSTM model offers predictions closer to the actual values. This indicates its effectiveness in capturing and predicting subtle variations in eye movement patterns, making it a valuable tool for accurately understanding and analyzing rapid changes in gaze behavior.

The error statistics of LSTM and ARIMA models are compared based on the performance metrics like mean absolute error (MAE), residual sum of squares (RSS), mean squared error (MSE), root mean squared error (RMSE), mean absolute percentage error (MAPE), mean error (ME) and mean percentage error (MPE) and the results are summarized in the Tables 2 and 3. This table provides a comprehensive comparison of the models and helps to conclude which is better suited for the given data.

After analyzing the consolidated error statistics of each feature based on each performance metric, it was observed that the values of the LSTM model were lower than those of the ARIMA models. The lower values for all of these evaluation measures in the LSTM model compared to the ARIMA model suggest that the LSTM model is better suited to capture the patterns in the data and provide more accurate predictions. After analyzing the error statistics of both models based on different performance metrics, it can be concluded that the LSTM model outperforms the ARIMA model in terms of accuracy and predictive power. The LSTM model is well-suited for capturing complex patterns and long-term dependencies in time series data, which could explain its better performance than the ARIMA model.

In the context of the PredictEYE model, the LSTM model was likely chosen for its ability to handle sequential data and capture long-term dependencies. While a powerful tool for time series forecasting, the ARIMA model may not be suitable for this task due to the non-linear and complex nature of the eye gaze sequence data. Therefore, the LSTM model will likely provide better accuracy and performance than the ARIMA model for predicting eye gaze sequences in the PredictEYE model [91].

2) PERFORMANCE EVALUATION OF PREDICTEYE IN ESTIMATING THE MENTAL STATE

This study evaluated six algorithms, namely Logistic Regression, Decision Tree, Random Forest, SVM, KNN, and Naive Bayes, for their performance in predicting the participants' mental states based on the predictions of ARIMA and LSTM models. After thoroughly analyzing and comparing the implemented machine-learning models, it was observed that the machine learning algorithms were performing better with LSTM predictions than with the ARIMA model.

The classification algorithms' performance was evaluated using various metrics, including accuracy, precision, recall, F1 score, and ROC curve, to determine the best-performing algorithm. The accuracy of the PredictEYE model refers to the percentage of correctly predicted mental states (calm or stressful) in the dataset. Precision is the percentage of correctly predicted instances of a specific mental state out of all the instances predicted as that mental state. Recall is the percentage of correctly predicted instances of a specific mental state out of all the actual instances of that mental state in the dataset. The F1 score is the harmonic mean of precision and recall for each mental state.

Figures 9 and 10 show the results compared across various machine learning algorithms based on the time series predictions of LSTM and ARIMA models. Prediction of mental state based on LSTM predictions showed better accuracy than predictions based on ARIMA. The analysis of the performance metrics results shows that all the classification algorithms performed better with the LSTM model than with ARIMA-based predictions.

Table 4 summarizes the performance metrics obtained from various models compared to PredictEYE. Among these models, PredictEYE, a combination of LSTM and Random Forest, achieved the highest accuracy of 86.4% and the highest F1 Score of 86.3% for Participant 4. Despite not achieving better precision and recall scores, the Random Forest model demonstrated superior overall performance

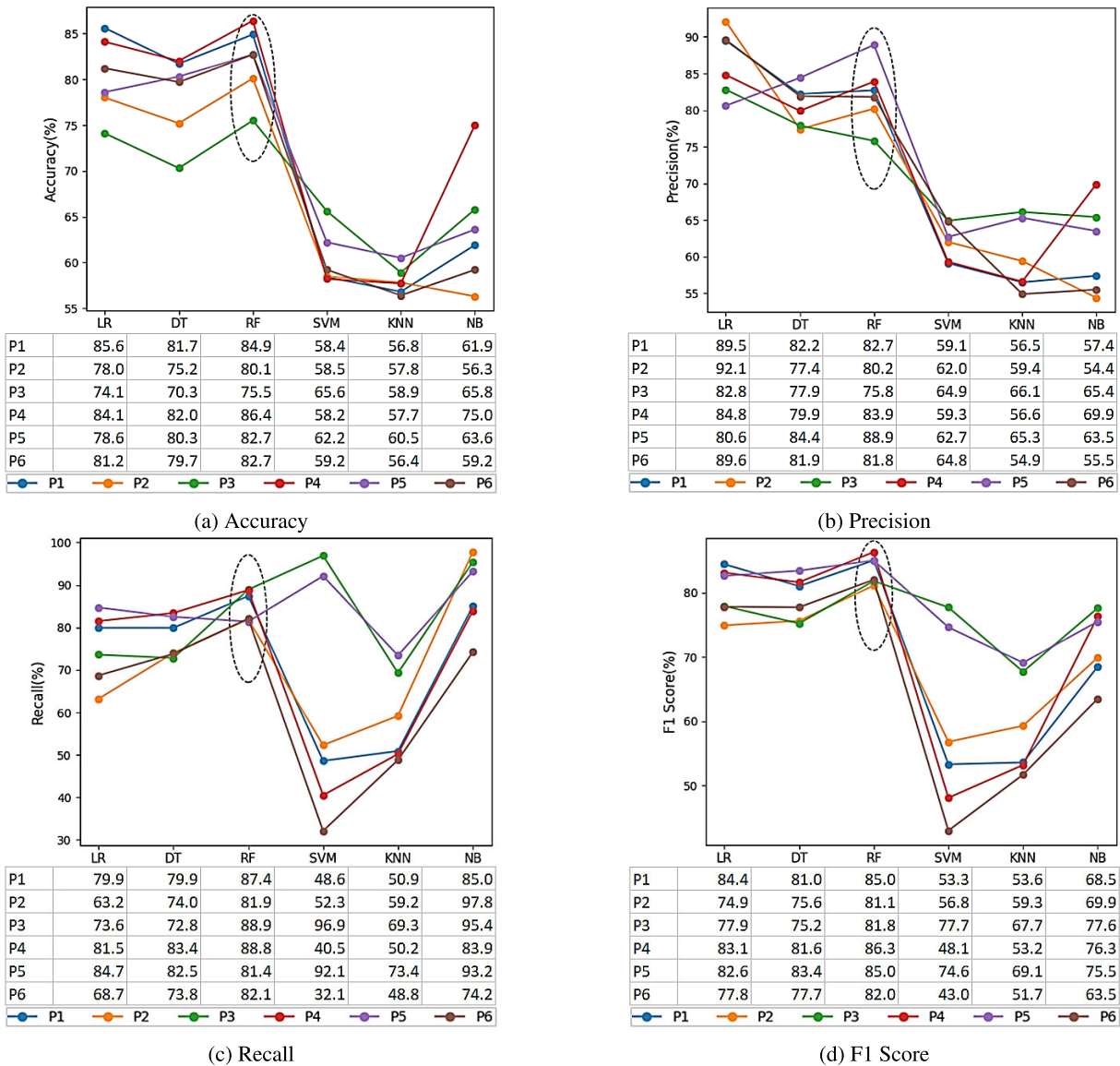


FIGURE 9. Analysis of classification model based on prediction with LSTM.

when considering all performance measures compared to the other models.

This finding highlights the importance of selecting the appropriate classification algorithm and evaluating the model’s performance using multiple metrics to understand its effectiveness comprehensively. Random Forest demonstrated a higher capability to accurately classify and predict the mental states based on the LSTM time series predictions, making it the most suitable choice for the PredictEYE model.

Figure 11 and 12 present the ROC curves and corresponding Area Under the Curve (AUC) values obtained from applying classification algorithms to the predictions generated by the LSTM and ARIMA models, respectively, for each participant. The ROC AUC values obtained from the classification algorithms applied to the LSTM-based predictions were consistently higher than those derived from the

ARIMA-based predictions. This indicates that the LSTM model’s predictions exhibited better discrimination between positive and negative instances across various classification thresholds. Among the classification algorithms evaluated, Random Forest consistently demonstrated higher performance in terms of AUC across most of the participants. The higher AUC values achieved by the classification algorithms applied to the LSTM-based predictions, along with the superior performance of Random Forest, indicate the strong predictive capabilities of the LSTM model in capturing meaningful patterns and features for classification tasks in this context.

In the analysis of classification models across all participants, the confidence intervals associated with the Receiver Operating Characteristic (ROC) curves, with a specific focus on Figures 11 and 12, were examined. These confidence

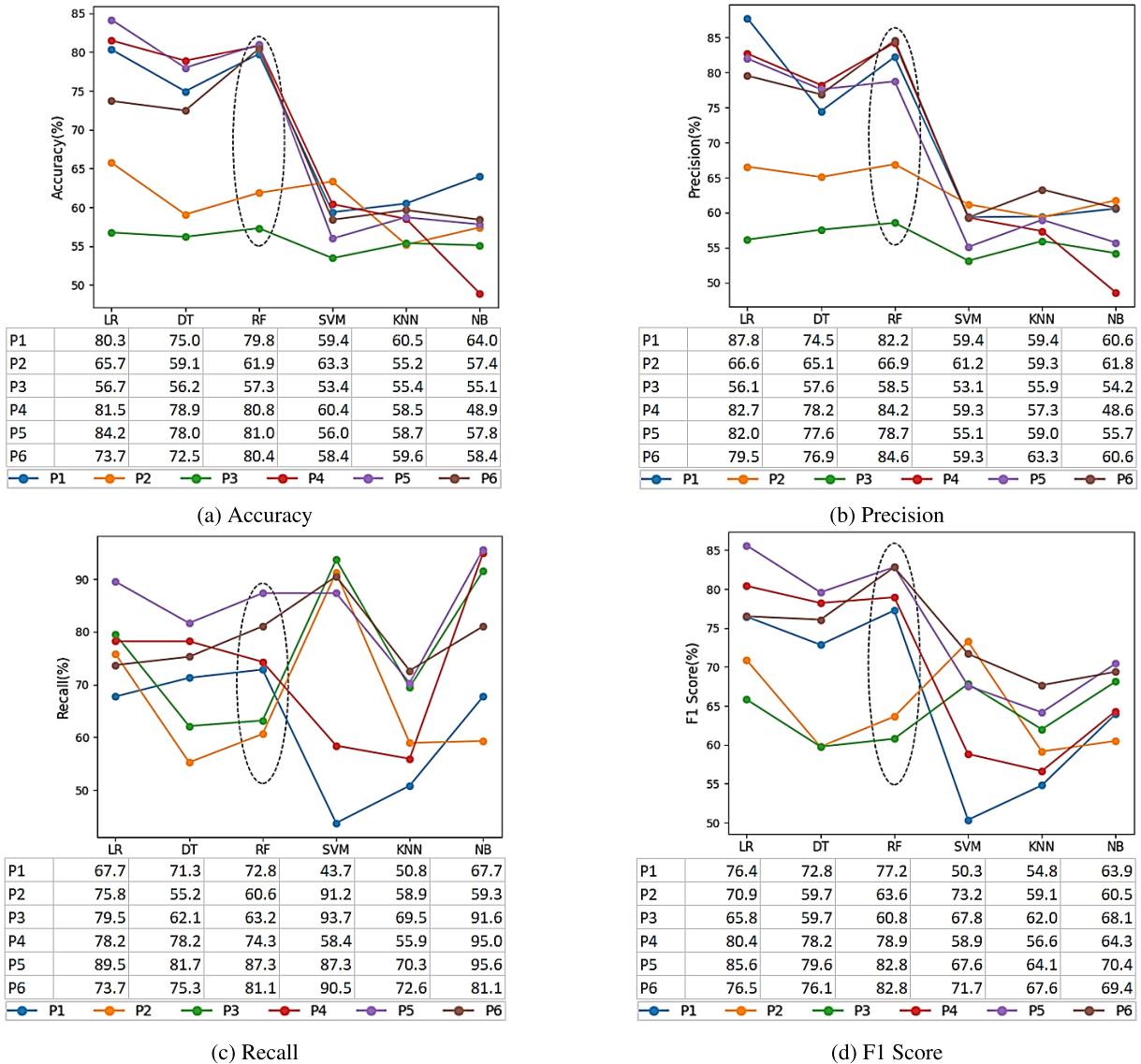


FIGURE 10. Analysis of classification model based on prediction with ARIMA.

intervals are valuable indicators of the precision and confidence level of our model estimates. Notably, for most participants, the Random Forest algorithm consistently exhibited narrow confidence intervals, signifying highly precise estimations and a strong degree of confidence that the true values fell within this range. To provide a more detailed illustration, the confidence interval details for participants P3 and P4 are shown in Table 5

Participant P4 emerged as a notable standout in our analysis, as the Random Forest algorithm achieved the highest classification accuracy for this individual. The accuracy, coupled with the narrow confidence intervals, indicates a high level of confidence in the model’s ability to accurately predict outcomes for P4. Conversely, for Participant P3, the Random Forest algorithm demonstrated the lowest classification accuracy among all participants. Despite this, the narrow confidence intervals for P3 suggest

that even in cases of lower accuracy, the model’s predictions are still made with high precision and confidence.

C. VALIDATION WITH GSR

Figure 13 presents the classification of mental states as either ‘calm’ or ‘stressful’ using Random Forest. This classification is based on the predicted sequences of eye tracking data collected at the completion of watching the calm and stressful video. The figure depicts the states determined by the Galvanic Skin Response (GSR) and the mental state predictions based on the predicted sequences of LSTM and ARIMA models. The visualization uses green to indicate the calm state and red to represent the stressful state.

When analyzing the mental state prediction based on eye-tracking features and comparing it to the GSR, the Random Forest algorithm could accurately predict the mental states of all participants using the predicted data sequence

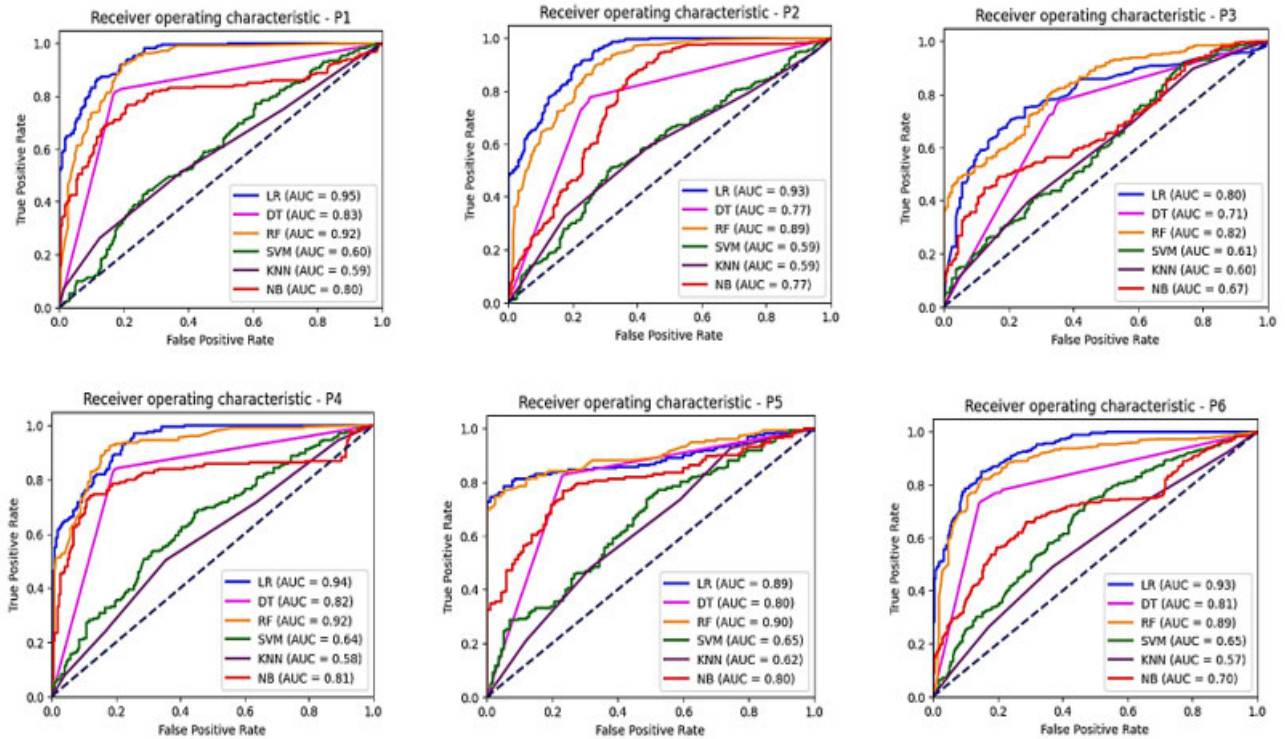


FIGURE 11. ROC curves based on classification algorithms applied after the prediction with LSTM for participants from P1 to P6.

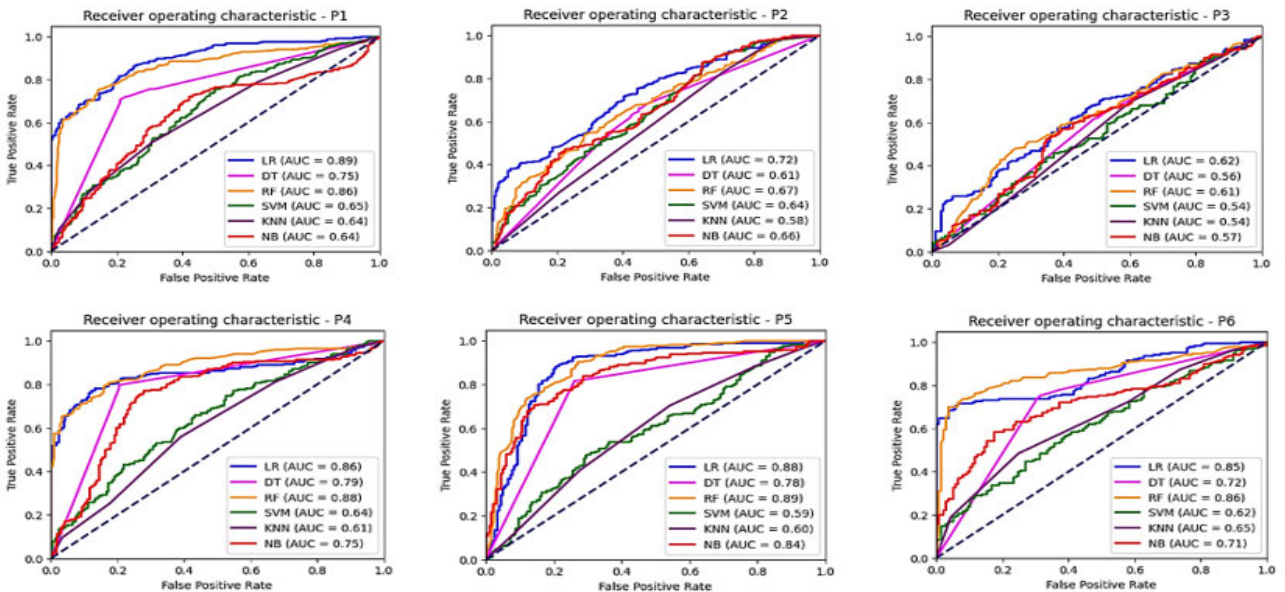


FIGURE 12. ROC curves based on classification algorithms applied after the prediction with ARIMA for participants from P1 to P6.

from the LSTM model. However, the algorithm failed to accurately predict the mental states of participants 3 and 5 when using the predicted sequence from the ARIMA model. These results suggest that the prediction of the LSTM model may influence the accuracy of the Random Forest algorithm in predicting mental states based on eye-tracking features.

Further analysis demonstrated that Participant 3 consistently exhibited a calm state according to the ground truth GSR during both videos. Random Forest successfully predicted the participant’s mental state at the end of watching the stressful video using the LSTM model’s predicted sequence. However, based on the ARIMA model’s predicted sequence, it yielded inaccurate predictions for the mental

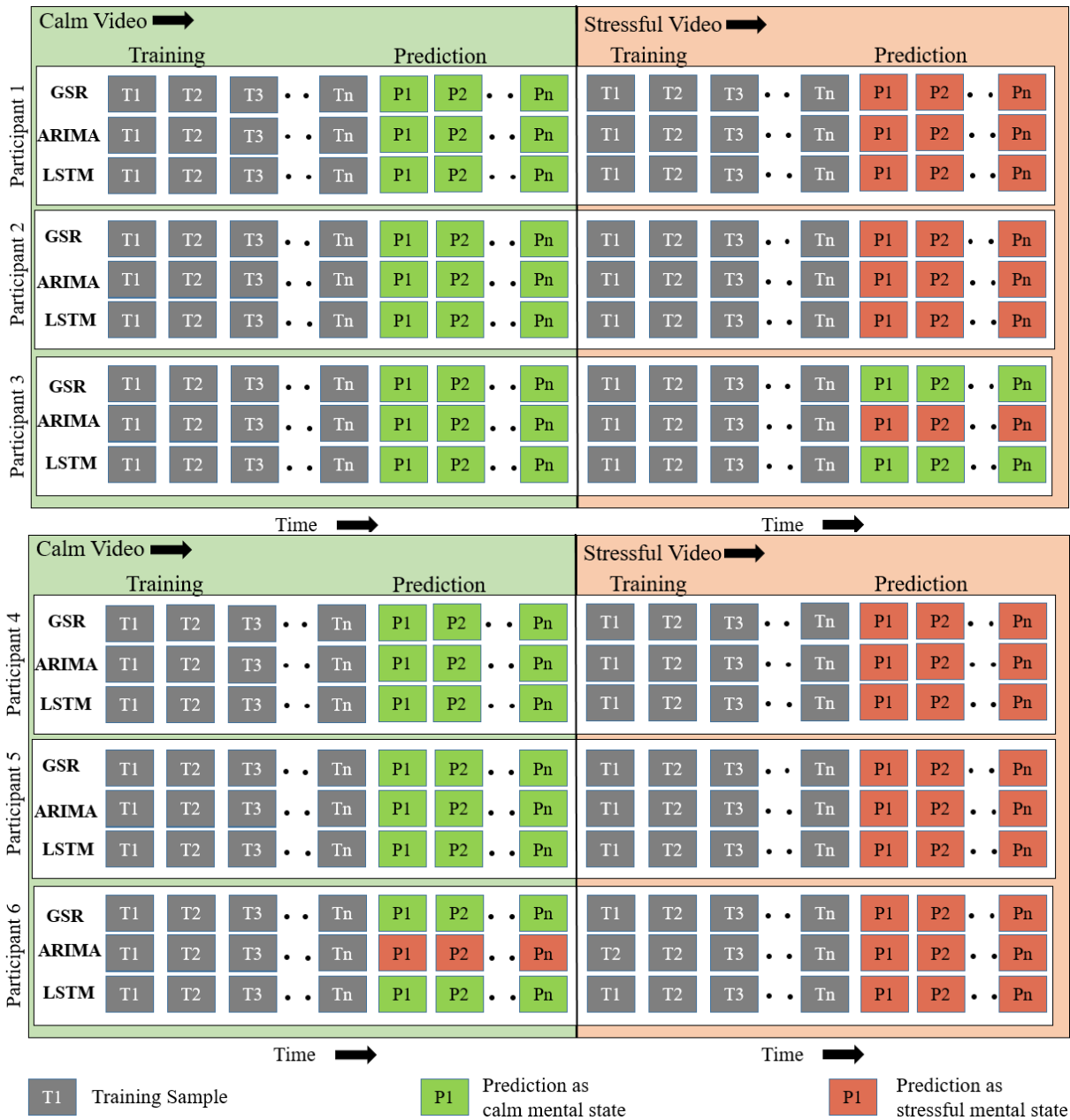


FIGURE 13. Forecasting of mental state based on eye tracking features considering GSR as ground truth.

state at the end of watching the stressful video. Additionally, the data underwent a Welch two-sample t-test, which was performed on the data samples obtained at the end of both calm and stressful video-watching periods. The results are presented in Table 1. Notably, certain features did not exhibit a significant difference at the conclusion of the stressful video, indicating that the participant remained calm throughout the observation. The findings indicate that the Random Forest algorithm’s performance in predicting mental states based on the predicted sequence generated by LSTM is superior to that of ARIMA’s predicted sequence.

Considering Participant 6, Random Forest could predict the mental state accurately based on the forecasted sequence

of LSTM for both videos. However, based on the predicted sequence of ARIMA data, the Random Forest algorithm classified the state as stressful for both videos, which is inconsistent with the GSR’s prediction. These results suggest that the performance of the PredictEYE model in predicting the mental state may depend on the specific features used and the algorithm utilized for data forecasting.

VI. DISCUSSION

PredictEYE demonstrates the effectiveness of predicting participants’ mental states based on their eye tracking data. The PredictEYE model utilizes LSTM to predict the future sequences of various eye tracking measures, such as pupil

TABLE 4. Performance evaluation of PredictEYE with other models.

Models	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
PredictEYE (LSTM+RF)	86.4	83.9	88.8	86.3
LSTM+LR	84.1	84.8	81.5	83.1
LSTM+DT	82	79.9	83.4	81.6
LSTM+SVM	58.2	59.3	40.5	48.1
LSTM+KNN	57.7	56.6	50.2	53.2
LSTM+NB	75	69.9	83.9	76.3
ARIMA+RF	80.8	84.2	74.3	78.9
ARIMA+LR	81.5	82.7	78.2	80.4
ARIMA+DT	78.9	78.2	78.2	78.2
ARIMA+SVM	60.4	59.3	58.4	58.9
ARIMA+KNN	58.5	57.3	55.9	56.6
ARIMA+NB	48.9	48.6	95	64.3

TABLE 5. Confidence Intervals associated with ROC curves.

Participant	Classification model	LSTM model with 95% of CI	ARIMA model with 95% of CI
P3	LR	75.96 to 83.63	57.44 to 67.45
	DT	66.79 to 75.46	51.27 to 61.48
	RF	78.63 to 85.92	56.44 to 66.44
	SVM	55.89 to 65.23	48.57 to 58.83
	KNN	55.37 to 64.72	48.38 to 58.64
	NB	62.86 to 71.82	52.21 to 62.38
P4	LR	91.44 to 96.03	82.14 to 88.89
	DT	78.74 to 85.97	75.58 to 83.34
	RF	89.68 to 94.76	85.25 to 91.41
	SVM	59.31 to 68.41	59.9 to 69.08
	KNN	53.41 to 62.76	55.89 to 65.27
	NB	77.61 to 85	71.25 to 79.52

diameter, blink rate, and fixation, while participants watch specific content on a screen. These eye tracking measures provide insights into the participants' mental states, whether they are moving towards a calm or stressful condition. For instance, when a participant is watching a stressful video, PredictEYE seeks to understand how their eye tracking metrics change over time and whether these changes lead them toward a calm or stressful state. The LSTM model plays a crucial role in this process by identifying patterns within each participant's eye tracking time series data and making predictions about their future eye movements and reactions. Random Forest algorithm is employed to collectively interpret the predicted future sequences of all eye tracking features. The Random Forest algorithm helps understand the participants' potential mental state to which they may be headed based on the predictions from the LSTM model. The insights gained from PredictEYE can be utilized to dynamically reorganize or skip the content being displayed to the participants, ensuring a more personalized and engaging experience based on their predicted mental states. This approach can be applied in various domains, such as mental health and stress management, to monitor and predict individuals' mental states in real-time based on their physiological data.

Using time series analysis on eye tracking data with high sampling frequency can provide several benefits in predicting

mental states with personalized models. Firstly, it allows for identifying unique patterns of gaze behavior associated with different mental states or disorders. Secondly, using time series analysis on eye-tracking data can help capture temporal dynamics and changes in mental states over time. By capturing changes in gaze behavior and mental states over time, personalized models can provide more accurate and timely predictions, allowing for more effective interventions and treatments.

In the PredictEYE model, eye tracking features play a significant role in predicting mental states. The model's utilization of LSTM-based time series analysis on eye tracking data enables it to capture unique gaze patterns, fixations, and eye movements associated with different mental states, such as calm or stressful. By leveraging these eye tracking features, the PredictEYE model can distinguish individual behavioral patterns, making its predictions more accurate and tailored to the specific mental states of each individual.

Compared to the ground truth Galvanic Skin Response (GSR), which predicts the mental state, the PredictEYE model's eye tracking features offer additional information. While GSR provides valuable physiological data related to mental state, eye tracking data goes beyond this by revealing what elements in the visual scene draw immediate attention and potentially influence the mental state. The PredictEYE model not only detects the mental state of an individual but also attributes the mental state to specific scenes using the information obtained from eye tracking. This feature allows for a more comprehensive understanding of the factors contributing to a person's mental state during video viewing. By combining eye tracking and mental state prediction, the PredictEYE model provides valuable insights into both conscious and subconscious responses, enhancing the accuracy of its predictions.

PredictEYE model has been compared with existing recent personalized and not personalized models [9], [24], [25], [26], [27], [93] in terms of stimulus, type of participants on which the study was performed, the features, algorithms used by the model, their achieved results, its analysis based on performance metrics, and type of the model as shown in Table 6. In eye tracking research, various stimuli, such as images, videos, tasks, and games, have been utilized to observe and analyze eye tracking measures. These measures typically include fixation, blink, saccade, and pupil diameter, employed in numerous studies to gain insights into different mental states. The PredictEYE model focused on using video stimuli as the input and extracted features based on fixation, blink, and pupil diameter to classify mental states as calm or stressful. Rather than comparing multiple users and attempting to understand the parameters responsible for mental state prediction, our approach aims to observe and comprehend individual patterns, considering the idiosyncratic nature of eye tracking data.

Numerous models have adopted machine learning and deep learning algorithms to classify mental states, attentional

TABLE 6. Comparison of PredictEYE with existing models.

Paper	Stimulus	Type of participant	Features	Algorithm	Result	Performance metric accuracy	Type of model (Personalized/other)
[31]	ST, AT, PT	–	PD, BR, FR,SR,ITD	CNN	Classification of attentional states	80%	Person-dependent and person-independent
[92]	SDS	Drivers	PD, BR, GD, VD,ED	CNN, LSTM	Classified the stress levels as low, medium and high	95.5%	Real time monitoring
[24]	V	ASD and TD children	FAA	DA	classification as ASD and TD	85.1%	Not personalized
[25]	V	-	PC	IP, ML	Detection of psychomotor impairment	–	Not Personalized
[16]	AW	with cancer without cancer	VM	CNN-LSTM	Identified mental health status	Hope- 93.8% , Anxiety-94.8%, Mental Wellbeing-95%	Self-proclaimed
[26]	V	with mental disorders	PD, FR, SR, FD, FF	CNN, DLSTM	Provides objective evaluation index of patients with mental disorders	–	Not Personalized
[27]	I	Normal	SSIM, E, C, HOC, PSD for EEG, and EOG-PDE, CGF, FV, RMSF for ET	GMM, DGNN	Classified emotions under the eight event stimuli	88.10%	Not Personalized
[9]	SEOE	Excavator operators from industry	BR, BD, PD, GP	TICC, SVM	Mental fatigue detection	85%	Not Personalized
[93]	RSSS	Surgical trainees	PD, FD, GE, PERCLOS	NB	Detection of perceived workload in robotic surgical tasks.	84.7%	Not Personalized
PredictEYE	V	Normal	PD, FD, BD, FDXY	LSTM, RF	Classified mental state as calm and stressful	86.4%	Personalized model Indication of scene responsible for mental state

ST – switch task, AT – alignment task, PT – pairs task, SDS – stressful driving situations, V – video, AW – art works, I – images, SEOE – simulated excavator operation experiment, RSSS – robotics skill simulation session, PD – pupillometric data, BR – blink rate, FR – fixation rate, SR – saccade rate, FD – fixation duration, BD – blink duration, ITD – imaging time series data, GD – gaze dispersion, VD – vehicle data, ED – environmental data, FAA – fixation at Area of Interest(eyes,mouth, body, hands, objects, background), PC – pupil centroid, VM – visual metrics, FF – fixation frequency, SSIM – self-similarity, E – Energy, C – Complexity, HOC – High order crossing, PSD – power spectral density, EOG-PDE – electrooculography power density estimation, CGF – center gravity frequency, FV – frequency variance, RMSF – root mean square frequency, ET – eye tracking, GP – gaze position, GE – gaze entropy, PERCLOS – Percentage of eyelid closure, FDXY – fixation dispersion on X and Y axis, DA – discriminant analysis, IP – image processing, ML – machine learning, NB – Naive Bayes, RF – Random Forest, CNN – Convolutional Neural Network, DLSTM – Deep LSTM, GMM – Gaussian Mixed Model, DGNN – Deep Gradient Neural Network, TICC – Toeplitz Inverse Covariance-Based Clustering, TD – typically developing

states, emotional states, and identify mental disorders, as well as detect perceived workload. Among these models, PredictEYE stands out with its unique approach, utilizing LSTM-based time series data prediction and random forest algorithm to predict mental states based on retrieved eye tracking data. The PredictEYE model falls under the personalized model category, intending to understand individual behavioral patterns while watching calm and stressful videos. By learning from these personalized patterns, the PredictEYE model predicts a person's mental state based on their unique eye tracking responses. PredictEYE is unique in its approach as it analyzes time series eye tracking data, thoroughly understands the unique eye tracking features of that person, and predicts their mental state and the specific scene responsible for it.

The performance of the PredictEYE model is compared to other models in terms of accuracy. The Random Forest model used in mental state prediction has shown promising results with a maximum accuracy of 86.4%. At the same time, it could not achieve such high accuracy for all the participants, but it is unique in its approach in detecting mental states. Collecting more data over a longer period can help better understand the unique patterns of an individual's mental state, leading to more accurate predictions and improved mental health outcomes.

A stressful mental state for Participant 1 might be attributed to scene 4, while for Participant 2, a different scene could be responsible for their mental state, as shown in Figure 14. The figure illustrates the mental state predictions of Participants 1, 2, and 3 while viewing a series of stressful scenes in a

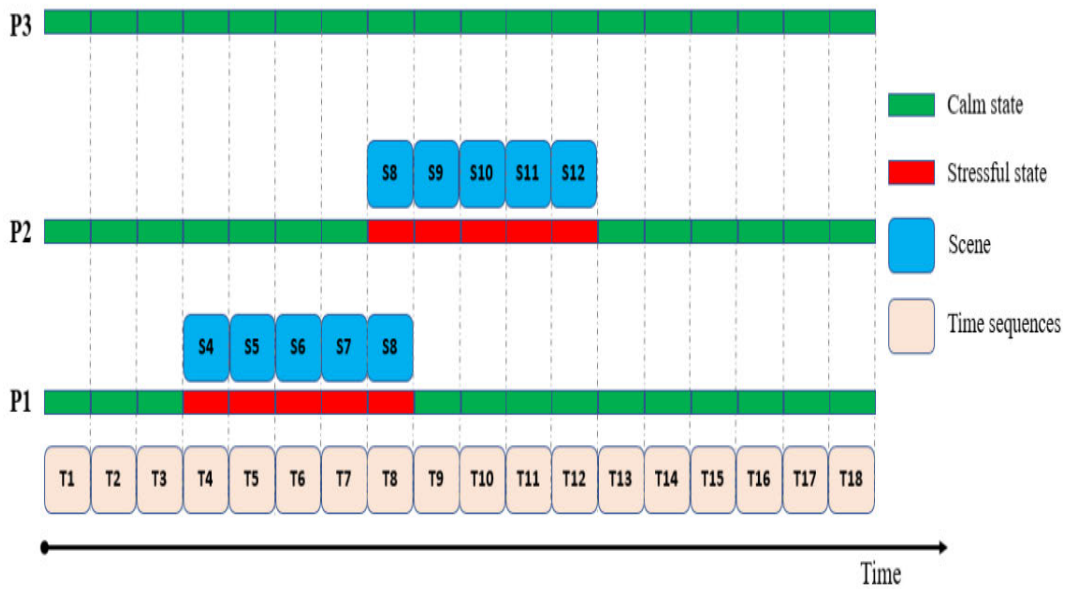


FIGURE 14. Sample scenes and gaze responsible for the mental state of participants.

video. The depicted time span ranges from T1 to T18, with each scene labeled S1 to S18. Participant 1 experienced a state of stress, and this was attributed to scene S4, which had a noticeable impact on their mental state. However, the same scene, S4, did not induce any changes in the mental states of Participants 2 and 3. Participant 2, initially in a calm state, transitioned into a state of stress due to scene S8. In contrast, Participant 3 remained consistently calm throughout the entire time span, with no observed alterations in their mental state caused by any scenes. These findings highlight the individual variability in how different participants respond to stressful stimuli and the unique triggers and reactions within their mental states. This capability to capture and differentiate individual responses is an exceptional characteristic of PredictEYE showcasing the diverse ways in which people perceive and react to stressful situations.

PredictEYE is a tool that focuses solely on normal individuals and aims to understand changes in their mental state by establishing a baseline period. During this baseline period, eye tracking measures are observed, and the model attempts to comprehend the trends and patterns of those measures to predict future mental states. This approach allows for the development of personalized, data-driven interventions to support individuals' mental well-being. By utilizing PredictEYE, individuals can gain insight into their mental state and make informed decisions about their mental health care.

One of the key advantages of using time series data analysis in the predictEYE model was the ability to develop personalized models for each participant. By analyzing the eye tracking data during both calm and stressful video viewing, the model was able to identify the underlying patterns in each participant's data and develop a

personalized model that could efficiently predict their mental state.

The PredictEYE model customizes its analysis of eye tracking data by utilizing LSTM-based time series models, which adapt to individual differences in a personalized manner. This personalized approach involves training the model on each person's specific eye tracking data, capturing their idiosyncratic patterns and responses. Instead of comparing data across multiple participants and treating them as a homogenous group, this personalized approach recognizes and respects the individuality of each person's cognitive and emotional processes.

The results demonstrated that the predictEYE model could accurately predict each participant's mental state based on their eye tracking data. The use of time series data analysis and personalized modeling in the predictEYE model could be applied to larger datasets in future studies to improve the accuracy and reliability of the model for mental state classification. By leveraging continuous monitoring of individuals, the PredictEYE model has the potential to identify patterns in an individual's mental state over time, providing valuable insights into the factors that contribute to stress or other mental health conditions. This could enable early interventions or treatments to be implemented before a condition becomes more severe.

During the development of the PredictEYE system, the challenge was to build a personalized model. Extensive literature surveys led to the discovery of suitable time series analysis methods for eye tracking data, enhancing personalization, and improving mental state prediction. Selecting the best predictive and classification model combination is crucial in developing accurate and efficient personalized models like the PredictEYE model, which could predict participants' mental states based on their eye tracking data.

By selecting the LSTM and the Random Forest models as the best models for predicting the mental state, the PredictEYE model could understand the trends in the time series data and classify participants' mental states into calm or stressed states with better performance.

VII. CONCLUSION

The PredictEYE system predicts a person's mental state and the scene responsible for it based on eye tracking data obtained while watching calm and stressful videos. The system employs an LSTM-based time series regression model to forecast future data sequences and a Random Forest algorithm-based classification model to forecast the mental state based on the future data sequences. The performance of the LSTM model was compared with that of an ARIMA model using error statistics obtained from performance measures, and the LSTM model was found to perform better. The performance of the Random Forest model was evaluated using various performance metrics and compared to other algorithms. Random Forest performed better than other algorithms in classifying individuals' mental states based on future data sequences. The PredictEYE model achieved a maximum accuracy of 86.4%, precision of 83.9%, recall of 88.8%, and an F1 score of 86.3% in predicting the mental state of a participant. The eye tracking features were found to have a significant role in predicting mental state, and the predictions based on these features were similar to the ground truth GSR. To ensure coordination between GSR and eye data, data was collected in the same machine, initiated both the data collection simultaneously, synchronized their timestamps, and used appropriate statistical analysis. These measures helped to minimize any lack of coordination and ensure the accuracy of the analysis. The PredictEYE model can incorporate various physiological signals to further improve the accuracy of mental state prediction.

The predictEYE model is a promising approach for predicting human mental states using eye tracking data. Using time series data analysis and personalized modeling in the predictEYE model could provide a more comprehensive understanding of the underlying patterns in the eye tracking data and enable more accurate predictions of the participants' mental state. The unique feature of PredictEYE to provide insights into the specific scene responsible for an individual's mental state makes it a valuable tool for understanding and predicting individuals' responses to different stimuli.

The PredictEYE model can be used as a screening tool for mental health disorders, such as anxiety and depression, by analyzing the eye-tracking data to predict a person's mental state. The model can also help monitor the effectiveness of treatment plans for mental health disorders. The PredictEYE model's adaptability and ability to integrate multiple physiological signals suggest that it has various potential applications in various domains, including healthcare and education.

The PredictEYE model can be adapted for webcam-based eye tracking, enabling continuous and non-invasive

monitoring of individuals' mental states and providing insights into stress levels, anxiety, or other emotional states over time as they work with the system. The accuracy of LSTM models can be improved by tuning their parameters and applying multivariate data analysis. Incorporating reinforcement learning in PredictEYE can improve the accuracy and personalization of mental state prediction by optimizing decision-making and adapting to changing mental state patterns over time. This can lead to better treatment and outcomes for individuals with mental health concerns. The adaptable and non-invasive nature of the PredictEYE model makes it a promising tool for continuously monitoring individuals' mental states in various applications, including healthcare, education, and employment.

ACKNOWLEDGMENT

The authors appreciate the help in collecting data for this research from Healthcare Global Enterprises Ltd. Cancer Centre, Bengaluru.

AUTHOR CONTRIBUTIONS

C. Jyotsna, J. Amudha, and Amritanshu Ram created theoretical formalism. C. Jyotsna and J. Amudha created the model's computational framework and design of the experiments. C. Jyotsna performed the data collection under the supervision of Amritanshu Ram. Amritanshu Ram provided support for the data collection. C. Jyotsna and J. Amudha did the analysis and interpretation of the results. C. Jyotsna prepared the draft manuscript. J. Amudha, Damiano Fruet, and Giandomenico Nollo reviewed the paper, and critical feedback from J. Amudha, Damiano Fruet, and Giandomenico Nollo helped shape the research, analysis, and manuscript.

DECLARATION OF CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

COMPLIANCE WITH ETHICAL STANDARDS

The research objective was explained to each participant, and their written consent was obtained. The Ethics Committee of HCG Cancer Hospital, Bengaluru, approved the research.

REFERENCES

- [1] R. Z. Zheng, *Cognitive Load Measurement and Application: A Theoretical Framework for Meaningful Research and Practice*. Evanston, IL, USA: Routledge, 2017.
- [2] A. Grisenti, G. Nollo, M. Dalprá, F. D. Natale, M. D. Cecco, A. Francesconi, A. Fornaser, P. Tomasin, N. Garau, and L. Guandalini, "Technological infrastructure supports new paradigm of care for healthy aging: The living lab Ausilia," in *Ambient Assisted Living: Italian Forum*. Cham, Switzerland: Springer, 2021, pp. 85–99.
- [3] R. Murthy, "National mental health survey of India 2015–2016," *Indian J. Psychiatry*, vol. 59, no. 1, p. 21, 2017.
- [4] M. Kivimäki, G. D. Batty, J. Pentti, M. J. Shipley, P. N. Sipilä, S. T. Nyberg, S. B. Suominen, T. Oksanen, S. Stenholm, M. Virtanen, M. G. Marmot, A. Singh-Manoux, E. J. Brunner, J. V. Lindbohm, J. E. Ferrie, and J. Vahtera, "Association between socioeconomic status and the development of mental and physical health conditions in adulthood: A multi-cohort study," *Lancet Public Health*, vol. 5, no. 3, pp. e140–e149, Mar. 2020.
- [5] N. Okkels, C. B. Kristiansen, and P. Munk-Jørgensen, *Mental Health and Illness in the City*. Cham, Switzerland: Springer, 2017.

- [6] J. Gruber, M. J. Prinstein, L. A. Clark, J. Rottenberg, J. S. Abramowitz, A. M. Albano, A. Aldao, J. L. Borelli, T. Chung, and J. Davila, "Mental health and clinical psychological science in the time of COVID-19: Challenges, opportunities, and a call to action," *Amer. Psychol.*, vol. 76, no. 3, p. 409, 2021.
- [7] S. Balters and M. Steinert, "Capturing emotion reactivity through physiology measurement as a foundation for affective engineering in engineering design science and engineering practices," *J. Intell. Manuf.*, vol. 28, no. 7, pp. 1585–1607, Oct. 2017.
- [8] Phasya. (2021). *Adaptive Software for Physiological and Cognitive States Monitoring*. Accessed: 2021. [Online]. Available: <https://www.phasya.com/en/technology>
- [9] J. Li, H. Li, W. Umer, H. Wang, X. Xing, S. Zhao, and J. Hou, "Identification and classification of construction equipment operators' mental fatigue using wearable eye-tracking technology," *Autom. Construct.*, vol. 109, Jan. 2020, Art. no. 103000.
- [10] A. P. Allen, P. J. Kennedy, S. Dockray, J. F. Cryan, T. G. Dinan, and G. Clarke, "The trier social stress test: Principles and practice," *Neurobiol. Stress*, vol. 6, pp. 113–126, Feb. 2017.
- [11] G. Acerbi, E. Rovini, S. Betti, A. Tirri, J. F. Rónai, A. Sirianni, J. Agrimi, L. Eusebi, and F. Cavallo, "A wearable system for stress detection through physiological data analysis," in *Italian Forum of Ambient Assisted Living*. Cham, Switzerland: Springer, 2016, pp. 31–50.
- [12] M. Parent, I. Albuquerque, A. Tiwari, R. Cassani, J.-F. Gagnon, D. Lafond, S. Tremblay, and T. H. Falk, "PASS: A multimodal database of physical activity and stress for mobile passive Body/ brain-computer interface research," *Frontiers Neurosci.*, vol. 14, p. 1274, Dec. 2020.
- [13] E. Garcia-Ceja, M. Riegler, T. Nordgreen, P. Jakobsen, K. J. Oedegaard, and J. Tøresen, "Mental health monitoring with multimodal sensing and machine learning: A survey," *Pervas. Mobile Comput.*, vol. 51, pp. 1–26, Dec. 2018.
- [14] A. N. Kurti, D. Davis, R. Redner, B. Jarvis, I. Zvorsky, D. R. Keith, H. Bolivar, T. J. White, P. Ripberger, C. Markeish, G. Atwood, and S. T. Higgins, "A review of the literature on remote monitoring technology in incentive-based interventions for health-related behavior change," *Transl. Issues Psychol. Sci.*, vol. 2, no. 2, pp. 128–152, Jun. 2016.
- [15] O. Postolache, D. J. Hemanth, R. Alexandre, D. Gupta, O. Geman, and A. Khanna, "Remote monitoring of physical rehabilitation of stroke patients using IoT and virtual reality," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 2, pp. 562–573, Feb. 2021.
- [16] S. B. Shafiei, Z. Lone, A. S. Elsayed, A. A. Hussein, and K. A. Guru, "Identifying mental health status using deep neural network trained by visual metrics," *Transl. Psychiatry*, vol. 10, no. 1, pp. 1–8, Dec. 2020.
- [17] P. Rathnayaka, N. Mills, D. Burnett, D. De Silva, D. Alahakoon, and R. Gray, "A mental health chatbot with cognitive skills for personalised behavioural activation and remote health monitoring," *Sensors*, vol. 22, no. 10, p. 3653, May 2022.
- [18] Y. Yang, H. Wang, R. Jiang, X. Guo, J. Cheng, and Y. Chen, "A review of IoT-enabled mobile healthcare: Technologies, challenges, and future trends," *IEEE Internet Things J.*, vol. 9, no. 12, pp. 9478–9502, Jun. 2022.
- [19] T. Sharon, "Self-tracking for health and the quantified self: Rearticulating autonomy, solidarity, and authenticity in an age of personalized healthcare," *Philosophy Technol.*, vol. 30, no. 1, pp. 93–121, Mar. 2017.
- [20] A. Obinikpo and B. Kantarci, "Big sensed data meets deep learning for smarter health care in smart cities," *J. Sensor Actuator Netw.*, vol. 6, no. 4, p. 26, Nov. 2017.
- [21] E. Serrano and J. Bajo, "Discovering hidden mental states in open multi-agent systems by leveraging multi-protocol regularities with machine learning," *Sensors*, vol. 20, no. 18, p. 5198, Sep. 2020.
- [22] F. Li, G. Xu, and S. Feng, "Eye tracking analytics for mental states assessment—A review," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2021, pp. 2266–2271.
- [23] Q. Wei, H. Cao, Y. Shi, X. Xu, and T. Li, "Machine learning based on eye-tracking data to identify autism spectrum disorder: A systematic review and meta-analysis," *J. Biomed. Informat.*, vol. 137, Jan. 2023, Art. no. 104254.
- [24] G. Wan, X. Kong, B. Sun, S. Yu, Y. Tu, J. Park, C. Lang, M. Koh, Z. Wei, Z. Feng, Y. Lin, and J. Kong, "Applying eye tracking to identify autism spectrum disorder in children," *J. Autism Develop. Disorders*, vol. 49, no. 1, pp. 209–215, Jan. 2019.
- [25] S. Adhikari and D. E. Stark, "Video-based eye tracking for neuropsychiatric assessment," *Ann. New York Acad. Sci.*, vol. 1387, no. 1, pp. 145–152, Jan. 2017.
- [26] Z. Xiang, W. Gao, T. Tao, L. Wang, C. Fan, and L. Xu, "Psychological test of patients with mental disorders based on eye movement data fusion algorithm," in *Proc. 4th Int. Conf. Inf. Syst. Comput. Aided Educ.*, Sep. 2021, pp. 2440–2445.
- [27] Q. Wu, N. Dey, F. Shi, R. G. Crespo, and R. S. Sherratt, "Emotion classification on eye-tracking and electroencephalograph fused signals employing deep gradient neural networks," *Appl. Soft Comput.*, vol. 110, Oct. 2021, Art. no. 107752.
- [28] A. Yadav, C. Jha, and A. Sharan, "Optimizing LSTM for time series prediction in Indian stock market," *Proc. Comput. Sci.*, vol. 167, pp. 2091–2100, Jan. 2020.
- [29] P. T. Yamak, L. Yujian, and P. K. Gadosey, "A comparison between ARIMA, LSTM, and GRU for time series forecasting," in *Proc. 2nd Int. Conf. Algorithms, Comput. Artif. Intell.*, Dec. 2019, pp. 49–55.
- [30] C. Jyotsna, J. Amudha, A. Ram, and G. Nollo, "IntelEye: An intelligent tool for the detection of stressful state based on eye gaze data while watching video," *Proc. Comput. Sci.*, vol. 218, pp. 1270–1279, Jan. 2023.
- [31] L.-M. Vortmann, J. Knychalla, S. Annerer-Walcher, M. Benedek, and F. Putze, "Imaging time series of eye tracking data to classify attentional states," *Frontiers Neurosci.*, vol. 15, p. 625, May 2021.
- [32] A. Balaskas, S. M. Schueller, A. L. Cox, and G. Doherty, "Ecological momentary interventions for mental health: A scoping review," *PLoS ONE*, vol. 16, no. 3, Mar. 2021, Art. no. e0248152.
- [33] J. Salminen, M. Nagpal, H. Kwak, J. An, S.-G. Jung, and B. J. Jansen, "Confusion prediction from eye-tracking data: Experiments with machine learning," in *Proc. 9th Int. Conf. Inf. Syst. Technol.*, Mar. 2019, pp. 1–9.
- [34] P. Kvtkn and T. Ramakrishnudu, "A novel method for detecting psychological stress at tweet level using neighborhood tweets," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 34, no. 9, pp. 6663–6680, Oct. 2022.
- [35] F. Al-shargie, T. B. Tang, N. Badruddin, and M. Kiguchi, "Towards multilevel mental stress assessment using SVM with ECOC: An EEG approach," *Med. Biol. Eng. Comput.*, vol. 56, no. 1, pp. 125–136, Jan. 2018.
- [36] C. Herbert, A. El Bolock, and S. Abdennadher, "How do you feel during the COVID-19 pandemic? A survey using psychological and linguistic self-report measures, and machine learning to investigate mental health, subjective experience, personality, and behaviour during the COVID-19 pandemic among university students," *BMC Psychol.*, vol. 9, no. 1, pp. 1–23, Dec. 2021.
- [37] R. Martinez-Maldonado, "'I spent more time with that team': Making spatial pedagogy visible using positioning sensors," in *Proc. 9th Int. Conf. Learn. Anal. Knowl.*, Mar. 2019, pp. 21–25.
- [38] C. Tsiourti, A. Weiss, K. Wac, and M. Vincze, "Multimodal integration of emotional signals from voice, body, and context: Effects of (In)congruence on emotion recognition and attitudes towards robots," *Int. J. Social Robot.*, vol. 11, no. 4, pp. 555–573, Aug. 2019.
- [39] M. J. Junokas, R. Lindgren, J. Kang, and J. W. Morphey, "Enhancing multimodal learning through personalized gesture recognition," *J. Comput. Assist. Learn.*, vol. 34, no. 4, pp. 350–357, Aug. 2018.
- [40] C. R. Henrie, R. Bodily, R. Larsen, and C. R. Graham, "Exploring the potential of LMS log data as a proxy measure of student engagement," *J. Comput. Higher Educ.*, vol. 30, no. 2, pp. 344–362, Aug. 2018.
- [41] E. Okur, N. Alyuz, S. Aslan, U. Genc, C. Tanriover, and A. A. Esme, "Behavioral engagement detection of students in the wild," in *Proc. Int. Conf. Artif. Intell. Educ.* Cham, Switzerland: Springer, 2017, pp. 250–261.
- [42] D. Rosen, A. Palatnik, and D. Abrahamson, "A better story: An embodied-design argument for generic manipulatives," in *Using Mobile Technologies in the Teaching and Learning of Mathematics*. Cham, Switzerland: Springer, 2018, pp. 189–211.
- [43] D. Di Mitri, M. Scheffel, H. Drachler, D. Börner, S. Ternier, and M. Specht, "Learning pulse: A machine learning approach for predicting performance in self-regulated learning using multimodal data," in *Proc. 7th Int. Learn. Anal. Knowl. Conf.*, Mar. 2017, pp. 188–197.
- [44] M. N. Giannakos, K. Sharma, I. O. Pappas, V. Kostakos, and E. Velloso, "Multimodal data as a means to understand the learning experience," *Int. J. Inf. Manage.*, vol. 48, pp. 108–119, Oct. 2019.
- [45] K. Sharma, P. Dillenbourg, and M. Giannakos, "Stimuli-based gaze analytics to enhance motivation and learning in MOOCs," in *Proc. IEEE 19th Int. Conf. Adv. Learn. Technol. (ICALT)*, Jul. 2019, pp. 199–203.
- [46] B. Schneider, K. Sharma, S. Cuendet, G. Zufferey, P. Dillenbourg, and R. Pea, "Leveraging mobile eye-trackers to capture joint visual attention in co-located collaborative learning groups," *Int. J. Comput.-Supported Collaborative Learn.*, vol. 13, no. 3, pp. 241–261, Sep. 2018.

- [47] H. Kurniawan, A. V. Maslov, and M. Pechenizkiy, "Stress detection from speech and galvanic skin response signals," in *Proc. 26th IEEE Int. Symp. Comput.-Based Med. Syst.*, Jun. 2013, pp. 209–214.
- [48] A. Grünerbl, A. Muaremi, V. Osmani, G. Bahle, S. Öhler, G. Tröster, O. Mayora, C. Haring, and P. Lukowicz, "Smartphone-based recognition of states and state changes in bipolar disorder patients," *IEEE J. Biomed. Health Informat.*, vol. 19, no. 1, pp. 140–148, Jan. 2015.
- [49] M. Ulinskas, R. Damaševičius, R. Maskeliūnas, and M. Woźniak, "Recognition of human daytime fatigue using keystroke data," *Proc. Comput. Sci.*, vol. 130, pp. 947–952, Jan. 2018.
- [50] C. Jyotsna and J. Amudha, "Eye gaze as an indicator for stress level analysis in students," in *Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI)*, Sep. 2018, pp. 1588–1593.
- [51] C. Jyotsna, J. Amudha, R. Rao, and R. Nayar, "Intelligent gaze tracking approach for trail making test," *J. Intell. Fuzzy Syst.*, vol. 38, no. 5, pp. 6299–6310, May 2020.
- [52] S. Krishnan, J. Amudha, and S. Tejwani, "Gaze exploration index (GE i)-explainable detection model for glaucoma," *IEEE Access*, vol. 10, pp. 74334–74350, 2022.
- [53] R. Pernice, Y. Antonacci, M. Zanetti, A. Busacca, D. Marinazzo, L. Faes, and G. Nollo, "Multivariate correlation measures reveal structure and strength of brain-body physiological networks at rest and during mental stress," *Frontiers Neurosci.*, vol. 14, p. 1427, Feb. 2021.
- [54] D. Carneiro, J. C. Castillo, P. Novais, A. Fernández-Caballero, and J. Neves, "Multimodal behavioral analysis for non-invasive stress detection," *Expert Syst. Appl.*, vol. 39, no. 18, pp. 13376–13389, Dec. 2012.
- [55] L. Yang and S.-F. Qin, "A review of emotion recognition methods from keystroke, mouse, and touchscreen dynamics," *IEEE Access*, vol. 9, pp. 162197–162213, 2021.
- [56] S. Klinghammer, T. Voitsekhijska, N. Licciardello, K. Kim, C.-K. Baek, H. Cho, K.-J. Wolter, C. Kirschbaum, L. Baraban, and G. Cuniberti, "Nanosensor-based real-time monitoring of stress biomarkers in human saliva using a portable measurement system," *ACS Sensors*, vol. 5, no. 12, pp. 4081–4091, Dec. 2020.
- [57] A. Thomaneck, M. Vollstedt, and M. Schindler, "What can eye movements tell about students' interpretations of contextual graphs? A methodological study on the use of the eye-mind hypothesis in the domain of functions," *Frontiers Educ.*, vol. 7, Dec. 2022, Art. no. 1003740.
- [58] S. Afroze, M. R. Hossain, and M. M. Hoque, "DeepFocus: A visual focus of attention detection framework using deep learning in multi-object scenarios," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 34, no. 10, pp. 10109–10124, Nov. 2022.
- [59] M. Lech, A. Czyżewski, and M. T. Kucewicz, "CyberEye: New eye-tracking interfaces for assessment and modulation of cognitive functions beyond the brain," *Sensors*, vol. 21, no. 22, p. 7605, Nov. 2021.
- [60] C. K. Ramachandra and A. Joseph, "IEyeGASE: An intelligent eye gaze-based assessment system for deeper insights into learner performance," *Sensors*, vol. 21, no. 20, p. 6783, Oct. 2021.
- [61] S. Krishnan, J. Amudha, and S. Tejwani, "Gaze fusion-deep neural network model for glaucoma detection," in *Proc. Symp. Mach. Learn. Metaheuristics Algorithms, Appl.* Cham, Switzerland: Springer, 2020, pp. 42–53.
- [62] P. Sowmyasri, R. Ravalika, C. Jyotsna, and J. Amudha, "An online platform for diagnosis of children with reading disability," in *Proc. Int. Conf. Comput. Vis. Bio Inspired Comput.* Cham, Switzerland: Springer, 2019, pp. 645–655.
- [63] M. L. Pavani, A. B. Prakash, M. S. Koushik, J. Amudha, and C. Jyotsna, "Navigation through eye-tracking for human-computer interface," in *Information and Communication Technology for Intelligent Systems.* Cham, Switzerland: Springer, 2019, pp. 575–586.
- [64] S. Krishnan, J. Amudha, and S. Tejwani, "Intelligent-based decision support system for diagnosing glaucoma in primary eyecare centers using eye tracker," *J. Intell. Fuzzy Syst.*, vol. 41, no. 5, pp. 5235–5242, Nov. 2021.
- [65] A. F. Klaib, N. O. Alsrehin, W. Y. Melhem, H. O. Bashtawi, and A. A. Magableh, "Eye tracking algorithms, techniques, tools, and applications with an emphasis on machine learning and Internet of Things technologies," *Expert Syst. Appl.*, vol. 166, Mar. 2021, Art. no. 114037.
- [66] N. Nourbakhsh, F. Chen, Y. Wang, and R. A. Calvo, "Detecting users' cognitive load by galvanic skin response with affective interference," *ACM Trans. Interact. Intell. Syst.*, vol. 7, no. 3, pp. 1–20, Sep. 2017.
- [67] M. Behroozi and C. Parnin, "Can we predict stressful technical interview settings through eye-tracking?" in *Proc. Workshop Eye Movements Program.*, Jun. 2018, pp. 1–5.
- [68] C. Pedrini, L. Marotta, and A. Guazzini, "ASMR as idiosyncratic experience: Experimental evidence," *Int. J. Environ. Res. Public Health*, vol. 18, no. 21, p. 11459, Oct. 2021.
- [69] D. F. Fisher, R. Karsh, F. Breitenbach, and B. D. Barnette, "Eye movements and picture recognition: Contribution or embellishment," in *Eye Movements and Psychological Functions.* Evanston, IL, USA: Routledge, 2021, pp. 193–210.
- [70] Tobii. (2023). *Tobii Connect*. Accessed: 2023. [Online]. Available: https://connect.tobii.com/s/article/galvanic-skin-response-gsr?language=en_US
- [71] R. Markiewicz, A. Markiewicz-Gospodarek, and B. Dobrowolska, "Galvanic skin response features in psychiatry and mental disorders: A narrative review," *Int. J. Environ. Res. Public Health*, vol. 19, no. 20, p. 13428, Oct. 2022.
- [72] M. V. Villarejo, B. G. Zapirain, and A. M. Zorrilla, "A stress sensor based on galvanic skin response (GSR) controlled by ZigBee," *Sensors*, vol. 12, no. 5, pp. 6075–6101, May 2012.
- [73] A. P. Field and R. R. Wilcox, "Robust statistical methods: A primer for clinical psychology and experimental psychopathology researchers," *Behaviour Res. Therapy*, vol. 98, pp. 19–38, Nov. 2017.
- [74] P. Cohen, S. G. West, and L. S. Aiken, *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences.* London, U.K.: Psychology Press, 2014.
- [75] K. H. Abdulkareem, M. A. Mohammed, A. Salim, M. Arif, O. Geman, D. Gupta, and A. Khanna, "Realizing an effective COVID-19 diagnosis system based on machine learning and IoT in smart hospital environment," *IEEE Internet Things J.*, vol. 8, no. 21, pp. 15919–15928, Nov. 2021.
- [76] S. Tuarob, C. S. Tucker, S. Kumara, C. L. Giles, A. L. Pincus, D. E. Conroy, and N. Ram, "How are you feeling: A personalized methodology for predicting mental states from temporally observable physical and behavioral information," *J. Biomed. Informat.*, vol. 68, pp. 1–19, Apr. 2017.
- [77] S. Shastri, K. Singh, S. Kumar, P. Kour, and V. Mansotra, "Time series forecasting of COVID-19 using deep learning models: India-USA comparative case study," *Chaos, Solitons Fractals*, vol. 140, Nov. 2020, Art. no. 110227.
- [78] Y. Acikmese and S. E. Alptekin, "Prediction of stress levels with LSTM and passive mobile sensors," *Proc. Comput. Sci.*, vol. 159, pp. 658–667, Jan. 2019.
- [79] X. Zheng, S. Hasegawa, M.-T. Tran, K. Ota, and T. Unoki, "Estimation of learners' engagement using face and body features by transfer learning," in *Proc. Int. Conf. Hum.-Comput. Interact.* Cham, Switzerland: Springer, 2021, pp. 541–552.
- [80] T. Hulsen, S. S. Jamuar, A. R. Moody, J. H. Karnes, O. Varga, S. Hedensted, R. Spreafico, D. A. Hafler, and E. F. McKinney, "From big data to precision medicine," *Frontiers Med.*, vol. 6, p. 34, Mar. 2019.
- [81] Y. Yamada and M. Kobayashi, "Detecting mental fatigue from eye-tracking data gathered while watching video: Evaluation in younger and older adults," *Artif. Intell. Med.*, vol. 91, pp. 39–48, Sep. 2018.
- [82] D. Joseph. (2015). *The World's Most Relaxing Film*. Accessed: 2023. [Online]. Available: <https://www.youtube.com/watch?v=dkFNdABPhCO>
- [83] (2015). *The Tiger Movie*. Accessed: 2023. [Online]. Available: <https://www.youtube.com/watch?v=WOSUBdaMjKs>
- [84] D. D. Salvucci and J. H. Goldberg, "Identifying fixations and saccades in eye-tracking protocols," in *Proc. Symp. Eye Tracking Res. Appl. (ETRA)*, 2000, pp. 71–78.
- [85] S. Kaushik, A. Choudhury, P. K. Sheron, N. Dasgupta, S. Natarajan, L. A. Pickett, and V. Dutt, "AI in healthcare: Time-series forecasting using statistical, neural, and ensemble architectures," *Frontiers Big Data*, vol. 3, p. 4, Mar. 2020.
- [86] IMotions. (2020). *Galvanic Skin Response (GSR): The Complete Pocket Guide*. Accessed: 2023. [Online]. Available: <https://imotions.com/blog/learning/research-fundamentals/galvanic-skin-response/>
- [87] S. Uday, C. Jyotsna, and J. Amudha, "Detection of stress using wearable sensors in IoT platform," in *Proc. 2nd Int. Conf. Inventive Commun. Comput. Technol. (ICICCT)*, Apr. 2018, pp. 492–498.
- [88] K. G. Ranjan, D. S. Tripathy, B. R. Prusty, and D. Jena, "An improved sliding window prediction-based outlier detection and correction for volatile time-series," *Int. J. Numer. Model., Electron. Netw., Devices Fields*, vol. 34, no. 1, Jan. 2021, Art. no. e2816.

- [89] H. Yu, E. B. Klerman, R. W. Picard, and A. Sano, "Personalized wellbeing prediction using behavioral, physiological and weather data," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Informat. (BHI)*, May 2019, pp. 1–4.
- [90] D. Benvenuto, M. Giovanetti, L. Vassallo, S. Angeletti, and M. Ciccozzi, "Application of the ARIMA model on the COVID-2019 epidemic dataset," *Data Brief*, vol. 29, Apr. 2020, Art. no. 105340.
- [91] Y.-T. Tsan, D.-Y. Chen, P.-Y. Liu, E. Kristiani, K. L. P. Nguyen, and C.-T. Yang, "The prediction of influenza-like illness and respiratory disease using LSTM and ARIMA," *Int. J. Environ. Res. Public Health*, vol. 19, no. 3, p. 1858, Feb. 2022.
- [92] 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.
- [93] C. Wu, J. Cha, J. Sulek, T. Zhou, C. P. Sundaram, J. Wachs, and D. Yu, "Eye-tracking metrics predict perceived workload in robotic surgical skills training," *Hum. Factors, J. Hum. Factors Ergonom. Soc.*, vol. 62, no. 8, pp. 1365–1386, Dec. 2020.



C. JYOTSNA received the B.Tech. degree in computer science and engineering from the Cochin University of Science and Technology, India, in 2004, and the M.Tech. degree in embedded systems from Amrita Vishwa Vidyapeetham, Bengaluru, India, in 2013.

She is currently an Assistant Professor with the Department of Computer Science and Engineering, Amrita School of Computing, Amrita Vishwa Vidyapeetham, with an experience of more than 17 years. The project titled "Diagnosis of reading disability in dyslexic children using eye tracking" has been funded by the Karnataka State Council for Science and Technology under the 41st series of student project programs in the academic year, from 2017 to 2018. Her research interest includes eye-tracking technology and its data analysis, image processing, machine learning, and deep learning.



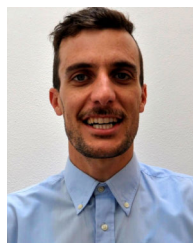
J. AMUDHA received the B.E. degree in electrical and electronics engineering from the Amrita Institute of Technology and Science, Coimbatore, India, in 1998, the M.E. degree in computer science and engineering, Anna University, Chennai, India, in 2002, and the Ph.D. degree in computer science and engineering from Amrita University, Coimbatore, in 2012.

She is currently a Professor with the Department of Computer Science, Amrita School of Engineering, Bengaluru. She is also a member of the UG and PG Board of Studies Member, the NAAC Accreditation, and an NBA Committee Member. She has filed a patent on an innovative technology in eye tracking research to automate the detection of optic discs for healthcare diagnostic systems. Her research interests include image processing, computer vision, soft computing, computational visual attention systems, cognitive vision, pattern recognition, and attention allocation in various fields, such as marketing and robotics.



AMRITANSHU RAM received the degree in microbiology and genetics and the master's and Ph.D. degrees in yoga research.

He is currently the Head of the Department of Complementary and Alternative Medicine, HCG Cancer Center, Bengaluru. He is also a Yoga Researcher with 18 years of experience working alongside cancer patients and researching the benefits of yoga on quality of life, psychology, and immune functions. With a passion for teaching and research, he drives the CAM Department, HCG Cancer Center, to be research-focused and has many ongoing collaborative studies with other renowned organizations. He has published more than 25 research articles in Indian and international journals in the field of oncology.



DAMIANO FRUET received the degree in biomedical engineering from Politecnico di Torino, in 2015.

He is currently carrying out research with the University of Trento. He primarily studies subject performance analysis and physiological data processing in the sports and medical fields. His research interests include physiological and environmental data acquisition, processing, and analysis employing feature extraction, and deep learning-based strategies.



GIANDOMENICO NOLLO (Member, IEEE) was born in Italy. He received the degree in physics from the University of Trento, Italy, and the Ph.D. degree in physiology from the University of Turin, Italy.

He is currently an Associate Professor of bioengineering with the Department of Industrial Engineering, University of Trento. He has conceived and led numerous national and international research projects aimed with the development and application of advanced health technologies in collaboration with academia, industries, and health facilities. The analysis of biological signals for modeling the interaction processes between physiological systems, digital health and the characterization of interface processes between implantable medical devices and tissues, functional performance, and safety of medical devices are the principal items of research. He is the author of more than 230 scientific publications (HI-Scopus of 40).

Dr. Nollo has been a member of several professional societies, including the IEEE. He is also a member of the scientific committee and the steering committee of the European Study Group of Cardiovascular Oscillations (ESGCO), the president of national and international workshops and conferences, the Co-Founder of the Italian Society for Health Technology Assessment (SIHTA), and the Deputy Vice President. He has received various awards for the work in biomedical engineering and medicine.

...

Open Access funding provided by 'Università degli Studi di Trento' within the CRUI CARE Agreement