Received April 12, 2021, accepted April 29, 2021, date of publication May 4, 2021, date of current version May 14, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3077358

Automatic and Intelligent Stressor Identification Based on Photoplethysmography Analysis

SAMI ELZEINY, (Student Member, IEEE), AND MARWA QARAQE, (Member, IEEE)
Information and Computing Technology Division, College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar

Corresponding author: Sami Elzeiny (selzeiny@hbku.edu.qa)

This work was supported by the Open Access funding from the Qatar National Library.

ABSTRACT A stressor is an external stimulus that an individual might consider demanding or threatening regarding his/her safety. Stressors cause the release of stress hormones and can be categorized into two categories: physiological and psychological. Wearable technology has been on the rise in the past decade, and the advancement of IoT will only further the benefits that wearable technology will bring. This work leverages the output of wearable technology to provide automatic stress and stressor identification model. In particular, this study proposes a novel algorithm that first detects instances of stress and then classifies the stressor type using photoplethysmography (PPG) data from wearable smartwatches. The time-series PPG data is transformed into 2D spatial images by first extracting the interbeat interval (IBI), and the blood volume pulse (BVP) features from the PPG, then transforming these data to 2D image data. The developed model successfully identified stress instances using IBI-BVP spatial domain images with an average accuracy of 98.10% with a convolutional neural network (CNN) and 99.18% using the average pixel intensity of these images with the extra trees classifier. It also successfully categorized stressor types into physical, cognitive, and social with an accuracy of 98.5% using the CNN and 96.45% using the extra trees classifier, outperforming state-of-the-art.

INDEX TERMS Stressor, stress state, convolution neural network, PPG sensor, interbeat interval, BVP signal, spatial domain, frequency domain, image processing, average pixel intensity.

I. INTRODUCTION

Stress is a kind of response in which the human body reacts to any threatening stimuli that is perceived (known as a stressor) in terms of behavioral, psychological, physiological, or physical responses. The inability to manage multiple cognitive loads can lead to a stressed state.

The event or the situation that happens to a person that triggers stress responses is called external stressors. In contrast, the inner thoughts or feelings, such as fears, beliefs, or lack of control that cause stress, known as internal stressors. Physical stressors cause mechanical pressure on bones, muscles, and skin, leading to pain and affecting work performance. Mental and social stressors may affect the way individuals respond to both chemical and physical stressors.

PPG is a simple, low cost and noninvasive optical technique used for heart rate monitoring. It utilizes a photodetector at the skin's surface and low-intensity infrared light, which is strongly absorbed by arterial blood. The volumetric changes in the blood are associated with cardiac activity and can be detected by PPG sensors as changes in the light intensity [1]. The PPG technique provides vital cardiovascular information and is widely used in several clinical applications such as measuring blood pressure, determining heart rate, diagnosing arterial disease, detecting microvascular blood flow, and blood oxygen saturation [1], [2].

In stressful conditions, the stress hormones, such as cortisol and adrenaline, influence heart rate, blood pressure, blood sugar level, and the ability to fight off illness. Long-term stress can increase the risk of heart stroke or attack and lead to depression. Furthermore, it can affect cognitive function, resulting in memory weakness. Therefore, early detection and prevention of mental stress are crucial to avoid psychological and physical health problems by preventing their causes or occurrences; accordingly, several approaches with different methods for detecting stress were proposed in the literature [3]–[6].

The pixel intensity value is an essential feature used in image classification. It is stored within pixels and equals a single value for the grayscale image and three values for

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

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

the colored image. The average pixel intensity (API) of the total image is the mean intensity value of a given image. Kamoen *et al.* applied the API method on the continuous-wave Doppler signal for grading the severity of mitral regurgitation (MR) caused by mitral valve prolapse (MVP). They found that the API method outperformed the other grading parameters such as vena contracta width and regurgitant volume by approximately 90% of all patients with MVP [7]. Pazinato *et al.* evaluated their proposed method for classifying tissues in ultrasound imagery at the pixel level using a dataset of five tissue types: calcium, blood, muscle, lipids, and fibrous. The classification accuracy of their method achieved 73%, which outperformed the threshold-based methods [8].

In this study, a stressor classification model is developed. First, a data collection effort was conducted to collect PPG data while inducing various stress types in university students. The proposed models derive both spatial and frequency images from IBI and BVP time-series data. Two system architectures are proposed. The first is based on deep learning for the model, and the second is based on machine learning (ML) methods. Deep learning-based systems require more training time and more powerful hardware with graphical processing units (GPUs) than machine learning systems due to the data amount that needs to be processed, the number of parameters that need to be tuned, and the complexity of the involved mathematical calculations. However, these systems are more robust, require less time in feature extraction, and are ideal for offline data processing. A machine learning-based approach allows for real-time detection and intervention as it is more lightweight than deep learning. In the ML model, pixel-based features that are discriminatory of stressor type are extracted for training and testing. The proposed work is ideal for the development of smart wearable technology that not only records vital signals but also monitors and alerts users of their health status.

This paper is structured as follows. Section II describes the data used in the study and the experimental protocol followed, and section III describes the proposed stress state and stressor type and image-based classification model. In section IV, the performance of the stress and stressor detection models are discussed. Finally, conclusions are discussed in Section VI.

II. DATA COLLECTION AND PREPROCESSING

This section presents details about the study protocol, participants, employed sensor, and self-reports of the two datasets used in this research. The first dataset is a result of an experiment conducted on six graduate students who were exposed to different types of stressors.

The ethical committee of the Qatar Biomedical Research Institute (QBRI), Hamad Bin Khalifa University (HBKU), approved this study on April 25, 2018, with IRB project number: 2018-016. All procedures performed in this study involving human participants were in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

The second dataset - the WESAD dataset [5]- is a publicly available multimodal dataset for wearable stress and affect detection. It includes physiological data of 15 participants to whom they were exposed to three affective conditions (baseline, amusement, stress).

A. PARTICIPANTS

In our study, the graduate students at our college were explicitly invited to participate in this study. A recruiting email was sent to recruit research participants. Exclusion criteria were mental disorders, depression, heavy smoking, chronic and cardiovascular diseases. To check whether the participants suffer from depression, they were asked to fill in two questionnaires: Beck's depression inventory (BD-II), which is a widely used questionnaire for detecting depression [9], and Depression Anxiety Stress Scales (DASS-21) to measure stress, anxiety, and depression [10]. All participants had a minimal level of depression and a normal level of stress and anxiety. In total, six graduate students were recruited with a mean age of 31±10.9 years (5 males, one female). In the WESAD dataset, 15 graduate students (12 males, three females) who had a mean age of 27.5±2.4 years participated in the study. Mental disorders, cardiovascular diseases, heavy smoking, and pregnancy were the exclusion criteria. Students were asked to avoid heavy exercise, caffeine, and tobacco before the experiment.

B. STUDY PROTOCOL

In our study, the experiment in Fig.1 was designed and conducted to simulate different types of stressors and detect stress

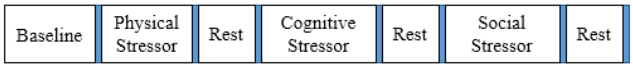

FIGURE 1. Study protocol to simulate different types of stressors. The blue boxes are the self-report stages.

based on physiological parameters. In order to prepare for the experiment, the subjects had to avoid caffeine, smoking, and doing challenging exercises on the experiment's day. Before the experiment, all subjects read and signed their informed consent form inclusion before participating in the study. Further, they filled out prequestionnaires to obtain their demographic information. The participants were equipped with an Empatica E4, which is a wristband wearable device designed for acquiring continuous real-time physiological signals in daily life. These signals are BVP, electrodermal activity (EDA), acceleration, and skin temperature. At the beginning of the study, a 20-minutes baseline measurement was initiated. Subsequently, the subjects were exposed to three stress stimuli: physical stressors (cold pressor test), cognitive stressors (mental arithmetic), and social stressors (giving an impromptu speech). These tasks were selected from two well-studied protocols: Trier Social Stress Test (TSST) and Socially Evaluated Cold Pressor Test (SECPT).

TSST consists of two main stressful tasks: giving a public speech and doing a mental arithmetic task [11]. In SECPT, the participants put their hands in ice water in front of other people [12]. In our version of the TSST and SECPT, the subjects were asked to put their dominant hand into ice water for only one minute and 30 seconds. Giles *et al.* found that the SECPT and TSST are the most robust effects on physiological measures [13]. The authors of [14] argued that the SECPT is a highly efficient tool to induce stress and activate major stress systems in a laboratory context. After the cold stressor, the participants had to solve the three levels of the mental arithmetic test within eight to ten minutes. Then, the participants were asked to deliver a five-minute speech in front of a panel. The participants had only 5 minutes to prepare their speech without using any supporting notes or slides during their presentation. Each stressful condition was followed by 5 minutes of relaxing mediation to de-excite the subjects and bring them back to the neutral state. At the end of the experiment, the E4 wristband was switched off. After each condition, the participant was asked to fill out a self-report questionnaire, which consists of the following questions:

- Valence level (with 5-point Likert type scale: 1-unpleasant, 2- Unsatisfied, 3-neutral, 4-pleased-5-pleasant).
- Arousal level (scale: 1-calm, 2-dull, 3-neutral 4-wide-awake,
  5-excited).
- Dominance levels (scale: 1-independent, 2- powerful, 3-neutral, powerlessness, dependent).
- Stress level (scale: 1-not at all, 2- slightly stressed, 3-moderately stressed, 4-very stressed, 5-extremely stressed).
- Stressor type.
- Stress effects on their body.
- Feeling type.
- Feeling level as used in the Positive and Negative Affect Schedule (PANAS) [15]).

The values of this subjective report express how the participant felt during each condition and can be utilized as data labels to train supervised machine learning algorithms to build personalized models. However, in this study, the study protocol was used as ground truth.

In the WESAD dataset, to collect the data, the subjects were equipped with chest- (RespiBan professional) and wrist- (Empatica E4) worn devices. RespiBan collects ECG, EMG, EDA, and TEMP signals with a sample rate of 700 Hz. At the same time, Empatica E4 records physiological data such as blood volume pulse (BVP/PPG - with a sampling rate =64 Hz), electrodermal activity (EDA - with a sampling rate = 4 Hz), skin temperature (TEMP - with a sampling rate = 4 Hz), and accelerometer data (ACC - with a sampling rate 32 Hz). The students were exposed to three affective conditions: baseline condition for 20 minutes while reading magazines, amusement for approximately 6 minutes and 32 seconds in which the participants watch eleven funny video clips, and stress condition by applying TSST protocol in which the subject delivers a 5-minute speech followed directly by a math task. A breathing exercise was used in the mediation stage after the amusement and the stressful situations to bring the subjects back to their neutral affective state.

C. DATA PREPROCESSING

The interbeat interval (abbreviated as IBI) is the time interval between the heart's beats and estimates the instantaneous heart rate. The IBI is a significant cardiac measure used to detect stress and provides an emotional state of the individual [16], [17]. The IBI sequence is obtained from the PPG signal processing by detecting peaks (beats) of the signal and computing the length of the interval between adjacent beats.

In our study, the IBI sequence was extracted from BVP/PPG data using Kubiso HRV premium software version 3.3.1 from Kubios Oy company in Kuopio, Finland [18]. Kubios is software used for heart rate variability (HRV) analysis; it includes algorithms and tools for artifact correction, pulse wave detection, noise handling, beat correction, and trend removal. Moreover, It calculates the time and frequency domain HRV parameter from electrocardiogram (ECG) and PPG signals. The BVP data obtained from E4 were imported into Kubios and used the automatic artifact correction method, which reduces the effect of possible artifacts by detecting missed, misaligned, and abnormal beats with high accuracy [19]. As the peaks of the PPG waveform are flat, the Kubios detects the rising edge of the pulse to calculate pulse-to-pulse intervals, as shown in Fig.2.

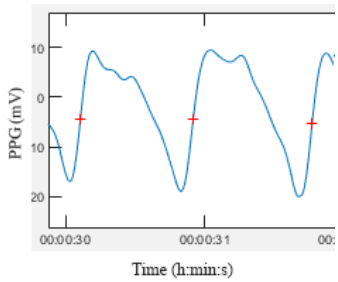

FIGURE 2. The Finding of the rising edge of a pulse to calculate pulse-to-pulse intervals (PPI or IBI).

III. METHOD

Time-series data comprise a sequence of data collected at different points in time. These data can be obtained by wearable sensing devices and present human health conditions. Time-series data may be transformed into 2D spatial images if a proper transformation function can be devised. In this study, PPG-IBI-based sequences and PPG-BVP-based values were converted to spatial images and were then transformed into frequency domain images.

A. IBI AND BVP IMAGE GENERATION

An image is represented by a 2D pixel matrix ($n \times m$), where each pixel in this image presents an integer element that

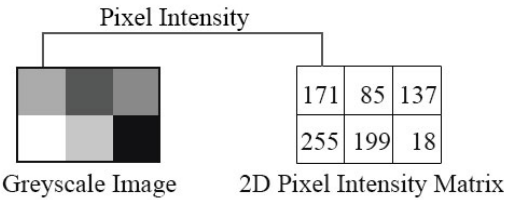

FIGURE 3. Representation of a grayscale image in the spatial domain.

corresponds to any one value, called the pixel intensity range from zero (black) to 255 (white) for an 8-bit image, as shown in Fig.3. 3D vectors of 2D matrices represent the RGB color images, and these matrices are a measurement of intensities of red, green, and blue components.

The state of these 2D matrices representing the intensity distribution of an image is called the spatial domain. In order to transform the time-series PPG data into spatial domain images, the interbeat intervals and the corresponding BVP values were first extracted from the collected PPG signal, as shown in Fig.4. The descriptive statistics such as the mean, standard deviation, minimum, maximum, and range were calculated for each of the two vectors. In the next step, 28 intervals were created by assigning a value of 1 to the interval in which the IBI value is located and a value of 0 to the remaining intervals. This method produces one matrix of size ($n \times m$), which is then utilized to generate IBI spatial domain images using a moving window of size 28 columns with an overlap of 27 columns (a one-column sliding window of size 28). The same procedure is followed to create another matrix for the BVP vector to generate BVP spatial domain images. Finally, both the IBI spatial images and BVP spatial images are combined to create new merged images that represent IBI and BVP values in the spatial domain, as shown in Fig.5.

The frequency-domain images were extracted by applying the Fourier transformation on the generated spatial domain images. The frequency in the image represents the rate of change of pixel values. In order to generate the frequency domain images, the fast Fourier transformation (FFT) was applied on the created spatial domain images. The frequency domain was investigated because the application of some filters is more computationally efficient in the frequency domain than the spatial domain. Moreover, frequency domain images provide different information than spatial domain images, which could be more useful in feature extraction for machine learning and deep learning models.

B. AVERAGE PIXEL INTENSITY CALCULATION AND EVALUATION

The pixel intensity value is a significant feature used for image classification; thus, to discriminate between stress and non-stress images as well as the different stressors, API is a noteworthy feature that can be derived. Thus, each spatial image was divided into four quadrants, and the API of each quadrant was calculated. These API values are used as input features to train and validate the candidate machine learning algorithms in order to classify both stress state and stressor type. Algorithm 1 was applied to calculate API for the images. Each spatial image was split into four quadrants; then, the average pixel intensity for the upper left, upper right, lower left, and lower right quadrants was calculated, as shown in Fig.6. As shown in Fig.7, there is a noticeable difference in pixel intensity value between the four quadrants for stressed and nonstressed IBI spatial images.

By comparing the average values of the lower and upper quadrants of the participants, it is found that they differ from one person to another as the amount of the stress response varies between them since the stress is subjective.

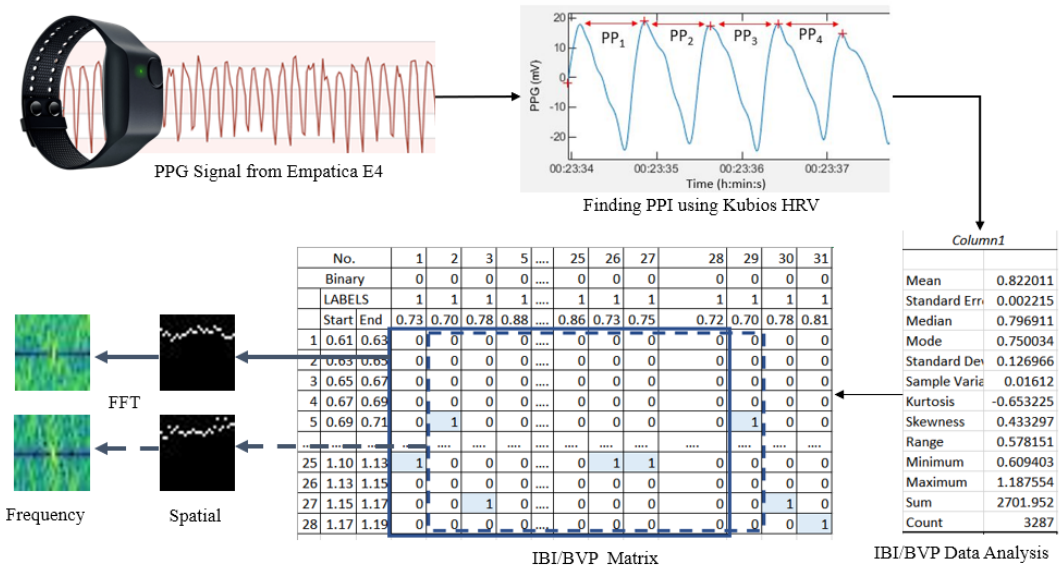

FIGURE 4. The Finding of the IBI values from PPG signals to convert IBI and BVP sequences matrices into spatial and frequency domain images.

FIGURE 5. The generation of IBI-BVP spatial domain images by combining IBI and BVP spatial images.

Algorithm 1 The Calculation of the API Values for the Four Quadrants of the Spatial Domain Image

Input: Consecutive IBI or BVP 28 × 28 matrix
Data: IBI_{data} or BVP_{data} or IBI_BVP_{data} as array
Output: Average Pixel Intensities for the four quadrants of the image

```
/* convert data matrix into image   */
Img ← imwrite(IBI_data);
/* split image into 4 quadrants     */
[rows,columns, dim] ← size(Img)
/* UL: Upper Left, UR: Upper Right,
   LL: Lower Left, LR: Lower Right  */
UL, UR, LL, LR ← imcrop(Img,[col,row,....]);
/* calculate average intensity for
   the 4 quadrants                  */
mean_UL ← mean2(UL)
API_UL ← [API_UL; k mean_UL]
mean_UR ← mean2(UR);
API_UR ← [API_UR; k mean_UR]
mean_LL ← mean2(LL);
API_LL ← [API_LL; k mean_LL]
mean_LR ← mean2(LR);
API_LR ← [API_LR; k mean_LR]
return API_UL,API_UR,API_LL,API_LR;
```

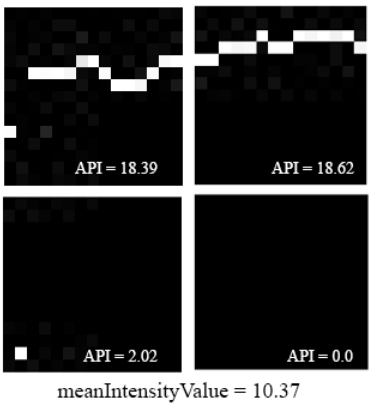

FIGURE 6. An example of calculating the API for the four quadrants of an IBI spatial image.

In order to facilitate the classification of stressor identification, the IBI spatial image was reinforced with BVP images.

Fig.8 and Fig.9 illustrate the API of stressor images for two different subjects. It is seen that the mean API values for the upper left and right for the cognitive stressor are greater than physical and social stressors for subject 1. While for subject 3, the values were lower for the cognitive stressor and greater for the physical stressor. This variation allows a machine-learning algorithm to discriminate the classes better and improves detection performance.

By combining the BVP spatial images with the IBI images, the level of variations between the API values for different stressors was kept the same as for the IBI images but with more comparative values, which will facilitate the work of the classifiers in distinguishing between the stressed and nonstressed states, as shown in Fig.8 and Fig.10.

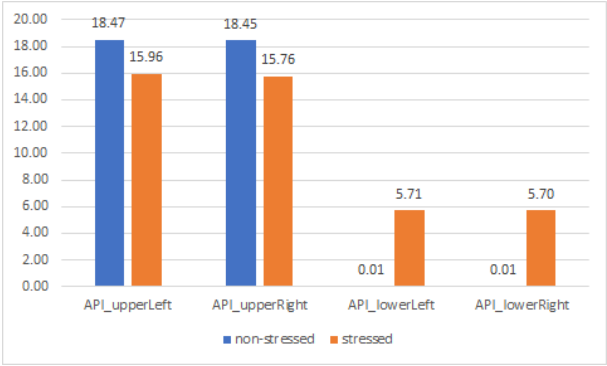

FIGURE 7. Subject 4 - Mean of API for the four quadrants of the IBI spatial domain images - stress states (nonstressed, stressed).

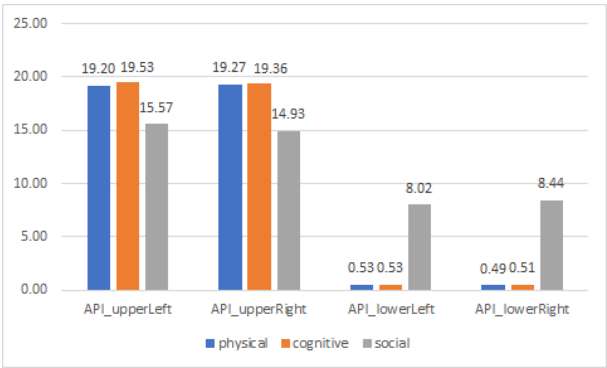

FIGURE 8. Subject 1 - Mean of the average pixel intensity for the four quadrants of IBI spatial images for the three stressors types.

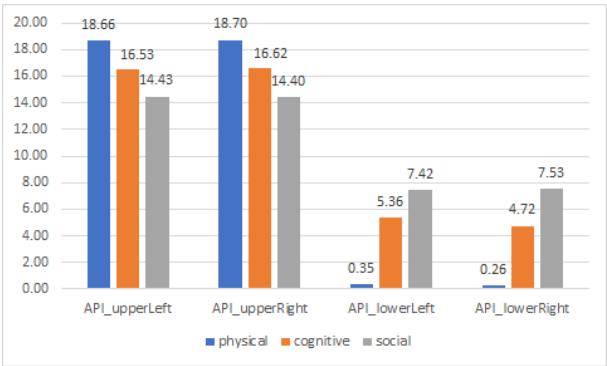

FIGURE 9. Subject 3 - Mean of the average pixel intensity for the four quadrants of IBI spatial images for the three stressors types.

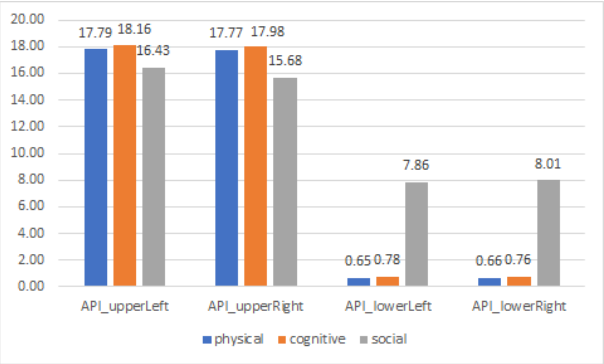

FIGURE 10. Subject 1 - Mean average pixel intensity for the four quadrants of IBI-BVP spatial images for the three stressors types.

TABLE 1. The size of the datasets used to train and validate deep learning and machine learning models.

Subject	Before Oversampling		After Oversampling	
	Baseline	Stressors	Baseline	Stressors
1	1,308	318	1,308	1,308
2	409	334	409	409
3	358	434	434	434
4	1,205	198	1,205	1,205
5	1,205	198	1,205	1,205
6	1,081	577	1,081	1,081
Total	**5,566**	**2059**	**5,642**	**5,642**

C. CLASSIFICATION MODELS

Stress is subjective and depends on a person's uniqueness (e.g., gender, age, genetics, personality, weight, coping ability, stress responses). Stress manifests differently in the body of each individual. In addition, various stressors manifest physiologically differently in each individual. Thus, personalized classification approaches must be used to optimize detection performance for each individual. In this section, two classification models are proposed and discussed. The first is based on deep learning, and the second is based on machine learning methods. A deep learning model is investigated as such models are more robust, do not require any feature extraction, and generally provide high-performance results. However, deep learning is heavier than many machine learning models and is not always feasible to be implemented in real-time applications or on wearable devices or smartphones. Thus, for real-time application for stress detection, machine learning models are more suitable; hence developing both models to support both offline and online detection was established. The performance of classification models is evaluated by measuring the accuracy of the training, validation, and testing, and the test sensitivity, specificity, and precision. The total number of generated spatial domain images is 11,824. These images were used to train, validate, and test the developed classification models. Table1 shows the dataset size per participant.

1) DEEP LEARNING-BASED CLASSIFICATION

A convolutional neural network (CNN or ConvNet) is a deep learning algorithm that takes an image as input and assigns weight and biases to the objects in the image to differentiate between these objects and classify an image under specific categories. The CNNs have attracted much attention in recent research, especially in image segmentation and classification, natural language processing (NLP), and face and object recognition due to their better performance [20]–[26]. The image is passed through a series of convolutional layers, pooling, fully connected layers, and softmax or sigmoid function classified with probability values between 0 and 1. Convolution layers with kernels or filters are mathematical operations used to extract features (feature map) from the image, indicating the strength and locations of the detected features in the image. An activation function such as a rectified linear unit (ReLU) is used to introduce nonlinearity in the CNN and transfer the gradient in training by backpropagation. The pooling layer, also called downsampling or subsampling, such as max pooling and average pooling, reduces the parameter numbers (dimensionality) of the convolved feature to decrease computational power in processing the images. The fully connected layer (FC layer) is used to flatten the output feature map matrix from convolution or pooling layers into a vector and drive the final classification decision. Activation functions, such as the sigmoid or softmax, are used to classify the outputs into a particular class.

In the CNN model, the input image of size 28×28 pixels passes through two convolutional layers to provide 32 and 64 feature maps with a convolution kernel of size 3×3 and the ReLU to apply an elementwise activation function. Two max-pooling layers were used to select the largest values on the convolved feature maps. The last layers are a fully connected layer with a depth of 128 and an output layer. The sigmoid function was used for the binary classification to classify the stress state into nonstressed and stressed, while the softmax function was for the multiclass classification to classify the stressor type into physical or cognitive or social, as shown in Figure 11. A five-fold cross-validation method was applied to validate our models. After each fold, the learned weight is initialized to the stored weight just after compiling the model.

2) MACHINE LEARNING-BASED CLASSIFICATION

In order to classify the stress state and stressor type, three different machine learning classifiers (extra trees, random forest, and support vector machine) are used. The features used to train and test these models are the API of the four quadrants of the IBI spatial images as discussed in section III-B. Cross-validation is a statistical method used to evaluate machine learning models and validate the model's performance on multiple data folds. In cross-validation, the dataset is split into ten folds. In the first iteration, the first fold is used to test the classification model, while the k-1 folds are used to train the model. In the second iteration, the 2nd fold is used to test the model, and the remaining data serve as the training

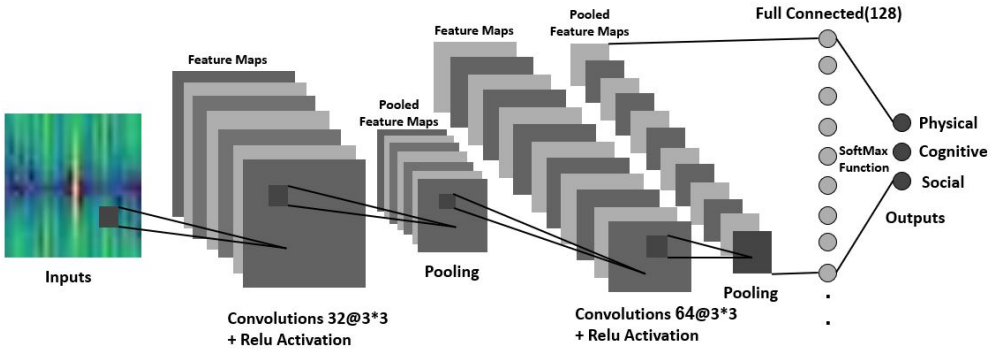

FIGURE 11. Stressors classification using CNN model.

set. This process is repeated until each fold has been utilized in both sets (training and testing datasets). A 10-fold cross-validation technique was adopted to ensure that each fold and every observation in the dataset has the chance to be used in both the training and test set. For the WESAD dataset, the generic model evaluation, the leave-one-subject out cross-validation (LOSO-CV) technique was used in which the classifier is tested using one subject dataset and trained on the remaining subjects datasets. In the training process, each classifier tries to learn the borderline of each class. The points nearby and on this borderline are more important for the classification and more possible to be misclassified than the points that exist far from the borderline. To handle imbalance between the classes, we applied the borderline-SMOTE technique, which is a new oversampling method to balance the class distribution in which the minority examples close to the borderline are oversampled [28].

Borderline-SMOTE finds k-nearest neighbors and draws a joining line between them to create new samples along the decision boundary between the two classes. It repeats these steps until the data are balanced. The performance of a model depends on the values of its hyperparameters. Examples for the hyperparameters for both extra trees and random forest are n_estimarots (the tress in the forest), max_depth (maximum number of levels in each decision tree), min_samples_leaf (minimum number of data points allowed in a leaf node), and min samples_split (minimum number of data points required in a node to split). For the support vector machine, the penalty parameter (C) adds a penalty to each misclassified data sample. The kernel is a mathematical function that takes input data and transforms it into the required form. An SVM has different kernel types, such as the Gaussian kernel, polynomial kernel, sigmoid kernel, and radial basis function (RBF). Gamma is a nonlinear hyperplane parameter; it determines how far the influence of the training samples reaches. In order to find the appropriate values of these hyperparameters, the GridSearchVC function was implemented. The function tries all the combinations of the predefined values for these hyperparameters and evaluates the model [27]. The machine learning algorithms used in this study to classify the stress state of the individual and the stressor type using the API of both IBI and BVP data are as follows:

- The random forest classifier (RF) or random decision forest is an ensemble tree-based learning algorithm that builds multiple decision trees and uses bootstrap replicas, and subsamples the input data with replacement. In the random forest, splitting nodes is based on the best split among a random subset of the features and selected at every node. It can be used in different applications, such as feature selection, image classification, and recommendation engines. It does not suffer from the overfitting problem because it takes the predictions' average of all decision trees participating in the classification or regression process, which cancels out the biases. For our RF models, the selected hyperparameters are: random_state =1, max_depth =13, n_estimators = 300, min_samples_split = 2, min_samples_leaf = 1, random_state =1, max_depth =13, n_estimators =300, min_samples_split = 2, min_samples_leaf = 1.
- The extra trees classifier or extremely randomized trees classifier is an ensemble learning method that was proposed by Geurts *et al.* in 2006 [29]. Extra trees build multiple decor-related decision trees in a forest and aggregate their votes to output the classification result. In the extra trees, the nodes are split based on random splits rather than best splits using random subsets of features. The extra trees classifier differs from the random forest classifier in constructing the decision trees in the forest and selecting the split point. By comparing the execution time and computational cost, the extra trees algorithm is faster than the random forest algorithm because instead of searching for the optimal split at each node, it selects it randomly. For our extra trees models, the selected hyperparameters are: max_depth = 25, min_samples_leaf = 1, min_samples_split = 2, n_estimators= 1200.
- The support vector machine (SVM) is a supervised machine learning model used for classification, pattern recognition, outlier detection, and regression. It finds a

hyperplane or decision boundaries in an N-dimensional space that best maximized the distance from the nearest data point of all the classes, where N is the number of features. For our SVM models, the selected hyperparameters are: C=1000, gamma= 0.001, kernel=rbf.

IV. PERFORMANCE OF DETECTION MODELS

In order to evaluate the performance of the created models, four metrics were calculated using the equation (1) - (4).

Accuracy presents the ratio of the correct classifications among the total classifications. The accuracy is calculated by the following equation, where the numerator represents correct classifications, while in the denominator are all classifications made by the model:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

where TP is the true positives, TN is the true negatives, FP is the false positives, and FN is the false negatives.

Sensitivity or recall (true positive rate): presents the fraction of the true positives divided by the sum of true positives plus false negatives. It measures the test capability to classify the stress state of an individual as stressed correctly. Sensitivity is calculated as follows:

$$Sensitivity = \frac{TP}{FN + TP} \tag{2}$$

The specificity (true negative rate): is the fraction of true negative divided by the sum of a true negative plus false positive. It measures the test capability to classify the stress state of an individual as nonstressed correctly. The formula used to calculate specificity is as follows:

$$Specificity = \frac{TN}{TN + FP} \tag{3}$$

Precision is the ratio of true positives to total classified positive. It measures the classifier's capability to classify positive out of all positives correctly. Precision is calculated as follows:

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

A. PERFORMANCE OF DEEP LEARNING MODELS

As shown in Table 6, the performance of the person-specific models that classify the stress state of the individual using IBI spatial domain images achieved higher accuracy than frequency domain images for subjects 1, 3, 4, and 6, while classifying stress using frequency domain images achieved more for subjects 2 and 5. The total average accuracy for the test dataset using spatial images was approximately 98.43%, while the frequency domain images were 97.18%. Adding more information to the IBI images by combining them with the BVP images enhanced the model's accuracy as the percentage increased to 100% for subjects 1, 3, and 6 for spatial images, and increased to 98.73%, 100%, and 98.65% for subjects 3, 5, and 6, respectively. For the stressor classification models, combining IBI and BVP helped the CNN models to achieve a high percentage for accuracy when classifying the stressor type using the frequency domain images. For the WESAD dataset, the average classification accuracy

TABLE 2. The performance measure for the classification models of stress state and stressor type using the API of the IBI-BVP spatial domain images with the ExtraTrees, SVM, and RF classifiers.

Subject	Images	Classifier	Test Accuracy (%)		
			Baseline	Stress	Amusement
2	1,570	CNN	100	95.96	95.96
		ExtraTrees	99.50	95.08	94.64
		SVM	94	85	93
		RF	100	97	97
3	1,318	CNN	86.55	86.55	100
		ExtraTrees	98.41	96.82	98.41
		SVM	97	88	92
		RF	98	95	97
4	1,742	CNN	100	100	100
		ExtraTrees	99.55	94.69	95.13
		SVM	99	81	80
		RF	99	89	90
5	1,162	CNN	100	93.99	93.99
		ExtraTrees	93.26	92.30	91.34
		SVM	90	75	75
		RF	98	90	92
6	1,466	CNN	100	100	100
		ExtraTrees	98.80	92.81	94.01
		SVM	96	77	81
		RF	99	87	86
7	1,908	CNN	91.22	99.27	90.49
		ExtraTrees	98.86	95.84	96.22
		SVM	96	77	78
		RF	99	87	86
8	1,447	CNN	97.27	87.37	90.1
		ExtraTrees	99.34	95.39	96.05
		SVM	99	72	74
		RF	99	89	90
9	1,684	CNN	94.95	87.37	92.42
		ExtraTrees	100	93	93
		SVM	94	68	73
		RF	100	84	84
10	2,089	CNN	100	100	100
		ExtraTrees	98.73	92.43	93.69
		SVM	92	81	89
		RF	98	90	92
11	1,833	CNN	83.74	88.62	95.12
		ExtraTrees	95.51	90.13	94.61
		SVM	67	73	72
		RF	95	82	87
13	1,739	CNN	90.94	86.88	95.94
		ExtraTrees	99.44	93.37	93.92
		SVM	93	73	77
		RF	99	92	92
14	2,118	CNN	95.12	100	95.12
		ExtraTrees	96.81	86.85	86.05
		SVM	92	72	67
		RF	97	82	80
15	1,806	CNN	100	93.92	93.92
		ExtraTrees	100	88.05	88.05
		SVM	100	70	69
		RF	100	85	85
16	1,587	CNN	100	98.71	98.71
		ExtraTrees	97.30	97.30	97.30
		SVM	78	77	82
		RF	97	93	96
17	1,443	CNN	100	100	100
		ExtraTrees	100	100	100
		SVM	98	100	98
		RF	100	100	100
Average		**CNN**	**95.99**	**94.58**	**96.12**
		ExtraTrees	**98.37**	**93.60**	**94.16**
		SVM	**90.33**	**76.22**	**77.89**
		RF	**99**	**92.67**	**93.67**

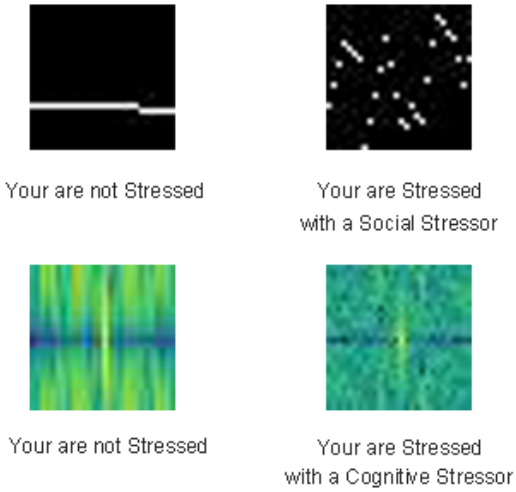

FIGURE 12. Stress state and stressor type classification results using the spatial and the frequency domain images.

TABLE 3. WESAD-The test accuracy measures for the person-specific affective state classification models using the IBI-BVP spatial domain images and the CNN, ExtraTrees, SVM, and RF classifiers.

Subject Out	Dataset Size	Test Accuracy (%)			
		Baseline	Stress	Amusement	Average
2	23,594	96.11	82.30	93.15	87.19
3	23,884	95.91	82.27	83.27	87.15
4	35,361	96.00	83.09	84.12	87.74
5	36,885	96.61	84.30	85.05	88.65
6	36,108	97.71	83.29	84.24	88.41
7	34,875	96.27	83.90	84.33	87.09
8	36,288	95.69	82.47	83.11	87.66
9	35,343	96.36	82.79	83.82	86.74
10	35,220	95.98	83.35	84.30	87.88
11	35,403	96.68	83.50	84.49	88.22
13	35,922	96.20	83.64	84.78	88.21
14	35,055	96.79	83.09	84.66	88.18
15	35,370	96.18	82.78	84.48	87.81
16	35,406	97.42	82.45	83.54	87.80
17	36,291	96.34	81.50	83.15	87.00
Average		**96.42**	**82.79**	**83.93**	**87.71**

achieved approximately 96.12% for the person-specific models and outperformed the used machine learning algorithms, as shown in Table 2.

In Fig. 12, some classification results use the CNN classification models to classify the stress state of the subjects and stressor type using IBI-BVP spatial and frequency domain images. These results represent the models' efficiency in classifying both stress state and stressor type.

B. PERFORMANCE OF MACHINE LEARNING MODELS

The performance of three machine learning algorithms (extra trees, SVM, and random forest) is discussed in this section.

1) PERFORMANCE OF EXTRA TREES MODELS

Table 5 illustrates the increase in performance when IBI and BVP are combined, as the average test accuracy achieved 99.18% when classifying the stressed state, and

TABLE 4. WESAD-The test accuracy measures for the generic classification models of stress and affective states using the IBI-BVP spatial domain images and LOSO-CV with the Extra Trees classifier.

Subject Out	Dataset Size	Test Accuracy (%)			
		Physical	Cognitive	Social	Average
1	3,648	96.21	94.15	97.93	96.10
2	4,473	93.55	92.43	98.87	94.95
3	3,264	95.40	92.33	96.93	94.89
4	4,248	92.92	91.15	98.23	94.10
5	4,473	92.99	89.91	96.91	93.27
6	3,834	92.15	89.21	97.05	92.80
Average		**93.87**	**91.53**	**97.65**	**94.35**

TABLE 5. The test accuracy measures for the generic classification models of stressor classification using the IBI-BVP spatial domain images and LOSO-CV with the Extra Trees classifier.

Subject	Classifier	Test Accuracy (%)			
		Stress State	Stressor Type		
			Physical	Cognitive	Social
1	ExtraTrees	100	94.5	94.5	100
	SVM	100	85	85	100
	RF	100	92	92	100
2	ExtraTrees	98.89	100	96	96
	SVM	99	100	96	96
	RF	99	100	96	96
3	ExtraTrees	100	93.38	93.38	100
	SVM	100	81	81	100
	RF	100	90	90	100
4	ExtraTrees	99.52	100	90.69	90.69
	SVM	100	93	79	86
	RF	100	100	93	93
5	ExtraTrees	100	100	92	92
	SVM	100	100	80	80
	RF	100	100	92	92
6	ExtraTrees	96.66	96	96	100
	SVM	92	86	86	100
	RF	97	96	96	100
Average	**ExtraTrees**	**99.18**	**97.31**	**93.76**	**96.45**
	SVM	**98.5**	**90.83**	**84.5**	**93.67**
	RF	**99.33**	**96.33**	**93.17**	**96.83**

97.31%,93.76%, 96.45% for physical, cognitive, and social stressors, respectively. By applying the LOSO-CV technique, the generic stressor type classification model achieved an average accuracy of 94.35% using the extra tree classifier and API value of the IBI-BVP spatial images, as shown in Table 5. As illustrated in Table 7, oversampling the minority classes using the borderline-SMOTE technique increased the training accuracy and the test accuracy, precision, sensitivity, F1-Score, and specificity for the cognitive and social stressors of the subject seven classification model.

In addition, it also increased the training accuracy, sensitivity, and F1 Score of the same stressors for subject 2.

For the WESAD dataset, the ExtraTrees classifier outperformed the other two algorithms (SVM, RF) in classifying affective states. As shown in Table 2, It achieved 98.37%, 93.60%, and 94.16% when classifying the baseline, stress, and amusement conditions, respectively.

Moreover, the three-class classification generic model (baseline vs. stress vs. amusement) achieved an average test accuracy of 87.71% using only BVP and IBI data as

TABLE 6. The average test accuracy for the classification models of stress state and stressor type using the IBI and IBI-BVP spatial and frequency domain images with CNN.

Subject	Image Type	Test Accuracy (%)									
		Stress State		Stressor Type							
				Physical		Cognitive		Social		Accuracy	
		IBI	IBI-BVP	IBI	IBI-BVP	IBI	IBI-BVP	IBI	IBI-BVP	IBI	IBI-BVP
1	Spatial	99.39	100	100	100	100	100	100	100	100	100
	Frequency	99.39	86.67	98.11	100	98.11	100	100	100	98	100
2	Spatial	100	100	100	100	100	91.3	100	91.3	100	91
	Frequency	100	96.32	100	100	100	86.96	100	86.96	100	100
3	Spatial	93.85	100	100	100	100	100	100	100	100	100
	Frequency	91.29	98.98	99.37	100	95.71	100	96.35	100	96	93
4	Spatial	100	100	100	100	100	100	100	100	100	100
	Frequency	98.74	98.73	92.86	100	92.86	100	100	100	93	100
5	Spatial	97.33	88.61	100	100	100	100	100	100	100	100
	Frequency	98.67	100	96.43	100	100	96.3	96.43	96.3	96	96
6	Spatial	100	100	100	100	100	100	100	100	100	100
	Frequency	95	98.65	100	97.37	100	97.37	100	100	100	97
Average	**Spatial**	**98.43**	**98.10**	**100**	**100**	**100**	**98.55**	**100**	**98.55**	**100**	**98.5**
	Frequency	**97.18**	**96.56**	**97.80**	**99.56**	**97.78**	**96.77**	**98.80**	**97.21**	**97.17**	**97.76**

TABLE 7. An example shows the impact of using oversampling with Borderline-SMOTE on the performance of the extra tree classifier.

Subject	Oversampling	Training 10-fold CV (%)	Measures	Stressors		
				Physical (%)	Cognitive (%)	Social (%)
2	Before	92.8,94.4,96,9.4,95.2,94.4 96,94.4,99.2,97.6	Accuracy	100	97.6	97.6
			Precision	100	90.62	100
			Sensitivity	100	100	70
			F1 Score	100	95.08	82.35
			Specificity	100	96.87	100
	After	96.8,97.3,95.1,98.2,98.6 95.5,91.9,97.7,95.9,97.3	Accuracy	100	97.32	97.32
			Precision	100	94.87	97.05
			Sensitivity	100	97.36	94.28
			F1 Score	100	96.10	95.65
			Specificity	100	97.28	98.70
7	Before	94.1,93.4,88.8,91.5,95.4 93.4,92.1,92.7,94.7,94.7	Accuracy	99.34	95.39	94.73
			Precision	100	88.63	83.33
			Sensitivity	98.90	95.12	75
			F1 Score	99.44	91.76	78.94
			Specificity	100	95.49	97.72
	After	95.8,96.6,93.9,94.3,95.8 95.8,94.3,94.7,94.3,95.4	Accuracy	98.86	95.84	96.22
			Precision	100	91.95	94.50
			Sensitivity	96.66	95.23	94.50
			F1 Score	98.30	93.56	94.50
			Specificity	100	96.13	97.12

shown in Table 4, which outperformed the accuracy achieved by [5] and [6] using multimodal data.

2) PERFORMANCE OF SVM MODELS

The performance of the SVM is evaluated in comparison with the other two classifiers, as shown in Table 5. The SVM classifier did not achieve higher accuracy than the RF and ExtraTrees classifiers in classifying stress state, except for the stress state classification using the API of the IBI-BVP, as it is achieved a classification accuracy of approximately 98.5% in classifying the stress state of the participant in our experiment. The same is true when classifying the stressor type; it did not outperform the other two classifiers in classifying the stressor type, as it achieved a classification accuracy of 90.83%, 84.5%, and 93.67% when using IBI-BVP. For the WESAD dataset, the SVM achieved only 90.33%, 76.22%, and 77.89% when classifying baseline, stress, and amusement conditions, respectively.

3) PERFORMANCE OF RANDOM FOREST MODELS

As illustrated in Table 5, the RF classifier achieved a higher accuracy when using the API of the IBI-BVP images to classify the stress state and the social stressor. The average test accuracy achieved by RF was approximately 99.33% in classifying the stress state and 96.83% in classifying the social stressor. For the WESAD dataset, the RF outperformed the ExtraTree classifier in classifying the baseline state, as it achieved an average accuracy of 99%, as shown in Table 2.

TABLE 8. The findings of this study compared with other studies.

Study	Features	Type	Classifier	Accuracy (%)
This Study	**IBI-BVP Spatial**	**CNN**	**Person-Specific**	**98.5**
	IBI-BVP Frequency	**CNN**	**Person-Specific**	**97.76**
	IBI-BVP API	**Extra Trees**	**Person-Specific**	**95.84**
	IBI-BVP API	**Extra Trees**	**Generic**	**94.5**
[3]	ECG-IBI Spatial	CNN	Generic	92.8
[4]	Raw ECG	CNN	Generic	90.19
[5]	ECG,EDA,EMG,RESP,TEMP	AdaBoost DT	Generic	80
[6]	ECG,EDA,EMG,RESP,TEMP	ANN	Generic	84.32%
[30]	PPG-time & frequency domain	DNN	Generic	91
[31]	HRV, EDA	Extra Trees,RF	Calibrated	95.2
[32]	PPG-time & frequency domain	SVM	Generic	92
[33]	PPG-time & frequency domain	RF	Generic	91
[34]	Body Video	SVM	Person-Specific	77
[35]	ECG,GSR,TEMP,RIP	SVM	Person-Specific	62

Based on the above discussions, the CNN and extra trees classifiers outperformed both classifiers (SVM and RF) in the person-specific classification model for stress state, stressor type, and affective states. Extra trees utilize the whole original sample and choose the cut points randomly, which reduces bias and variance, whereas RF subsamples the input data with replacement and chooses the optimum split.

C. MODEL PERFORMANCE COMPARISON

In Table 8, a list of previous approaches that were proposed to develop either generic or person-specific models to detect stress state, stress level, or affective states is shown. The performance of our models achieved high accuracy and comparative scores with the other models taking into account the different types of data that were used in our approach (spatial and frequency domain images extracted from PPG signal, and average pixel intensity values from both IBI and IBI-BVP spatial images), and stressor type classification. The results of both the deep learning and ML models indicate that spatial- and frequency-based PPG images hold immense stress-based characteristics that intelligent systems can exploit for the automatic detection of stress.

V. DISCUSSION

Stress is a body reaction to any threatening stimulus (stressor) that is perceived. Prolonged stress can seriously affect overall human well-being. It can lead to both physical and mental illness. Therefore, it is important to detect stress in the early stage to prevent health problems. Stress is subjective and is manifested differently from one person to another. Thus, there is a need to develop a method to detect the stress state of the individual and stressor type effectively to provide the user with the optimal strategies to increase his ability to cope with the stressful situations. The rapid growth in the wearable health monitoring technology market makes embedded PPG sensors in the wearable sensing devices a potential solution that can measure the volumetric variations of blood circulation influenced by the cardiac, vascular, and autonomic nervous systems, all affected by stress. In this article, a new PPG-based image analysis is proposed in which the IBI and the BVP are extracted from wearable-based PPG signals and converted to spatial domain images and transformed to the frequency domain. These generated images and the calculated average pixel intensity values from the spatial domain images are used to train, validate, and test deep learning and machine learning classification models to classify the individual's stress status into stressed and non-stressed, and the stressor type into physical, cognitive, and social. Combining IBI and BVP enhanced the models' accuracy helped the classification models to achieve a high percentage for accuracy when classifying the stressor type using the frequency domain images. Moreover, this combination enhanced the other accuracy measures such as precision, sensitivity, and specificity for stress state and stressor types classification models. The developed classification models were evaluated on two datasets, and the results show that our proposed methods successfully classify the stressor type and the participants' stress state with higher accuracy and outperformed state of the art. Our approach can be applied to monitor the individual's psychological wellbeing daily by classifying his stress state and stressor type using information extracted from the PPG signal. This approach can be utilized in several premises such as offices and educational institutes to detect when the employees and the students are experiencing stress. This will enable the managers and the teachers to intervene and handle stressful causes to increase the performance and reduce the stress of their staff.

VI. CONCLUSION

Everyone experiences stress at some point or another; however, if prolonged, stress can adversely affect individuals. This work proposes a new approach to build a person-specific model to detect the stress state and stressor type accurately, and the approach continuously monitors the perceived stress of the person in real life. The CNNs were employed to classify spatial and frequency domain images of the extracted IBI and BVP from a PPG sensor into stress states (stressed,

nonstressed) and stressor types (physical, cognitive, social). Deep learning-based systems need more powerful hardware and more processing time to train the model and tune the parameters due to the complex calculations and data size. Thus, a machine learning-based approach using the API was proposed to classify the stress state of the individual and the stressor type. Both proposed approaches achieved very high performance and outperformed the state-of-the-art approaches. The limitation of this study is mainly the size of the collected data. Future work includes further validating the performance of the models on a larger dataset by recruiting more graduate and undergraduate students with different age range to participate in our study and examining other wearable-based physiological signals such as EDA and skin temperature. However, the novelty of the work shows that PPG is an excellent data source when the proposed time series to image transformation is used. In addition, PPG sensors can be easily incorporated into wearable devices than ECG sensors.

REFERENCES

[1] M. Ghamari, "A review on wearable photoplethysmography sensors and their potential future applications in health care," *Int. J. Biosensors Bioelectron.*, vol. 4, no. 4, p. 195, 2018.

[2] T. Y. Abay and P. A. Kyriacou, "Photoplethysmography for blood volumes and oxygenation changes during intermittent vascular occlusions," *J. Clin. Monitor. Comput.*, vol. 32, no. 3, pp. 447–455, Jun. 2018, doi: 10.1007/s10877-017-0030-2.

[3] J. Huang, X. Luo, and X. Peng, "A novel classification method for a Driver's cognitive stress level by transferring interbeat intervals of the ECG signal to pictures," *Sensors*, vol. 20, no. 5, p. 1340, Feb. 2020.

[4] H.-M. Cho, H. Park, S.-Y. Dong, and I. Youn, "Ambulatory and laboratory stress detection based on raw electrocardiogram signals using a convolutional neural network," *Sensors*, vol. 19, no. 20, p. 4408, Oct. 2019.

[5] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, "Introducing WESAD, a multimodal dataset for wearable stress and affect detection," in *Proc. 20th ACM Int. Conf. Multimodal Interact.*, Oct. 2018, pp. 400–408.

[6] P. Bobade and M. Vani, "Stress detection with machine learning and deep learning using multimodal physiological data," in *Proc. 2nd Int. Conf. Inventive Res. Comput. Appl. (ICIRCA)*, Jul. 2020, pp. 51–57.

[7] V. Kamoen, M. El Haddad, T. De Backer, M. De Buyzere, and F. Timmermans, "The average pixel intensity method and outcome of mitral regurgitation in mitral valve prolapse," *J. Amer. Soc. Echocardiography*, vol. 33, no. 1, pp. 54–63, Jan. 2020.

[8] D. V. Pazinato, B. V. Stein, W. R. de Almeida, R. de O. Werneck, P. R. M. Junior, O. A. B. Penatti, R. D. S. Torres, F. H. Menezes, and A. Rocha, "Pixel-level tissue classification for ultrasound images," *IEEE J. Biomed. Health Informat.*, vol. 20, no. 1, pp. 256–267, Jan. 2016, doi: 10.1109/JBHI.2014.2386796.

[9] A. T. Beck, R. A. Steer, R. Ball, and W. F. Ranieri, "Comparison of beck depression inventories-IA and-II in psychiatric outpatients," *J. Personality Assessment*, vol. 67, no. 3, pp. 588–597, 1996.

[10] J. D. Henry and J. R. Crawford, "The short-form version of the depression anxiety stress scales (DASS-21): Construct validity and normative data in a large non-clinical sample," *Brit. J. Clin. Psychol.*, vol. 44, no. 2, pp. 227–239, Jun. 2005, doi: 10.1348/014466505X29657.

[11] C. Kirschbaum, K.-M. Pirke, and D. H. Hellhammer, "The 'trier social stress test'–a tool for investigating psychobiological stress responses in a laboratory setting," *Neuropsychobiology*, vol. 28, nos. 1–2, pp. 76–81, 1993, doi: 10.1159/000119004.

[12] N. Minkley, T. P. Schröder, O. T. Wolf, and W. H. Kirchner, "The socially evaluated cold-pressor test (SECPT) for groups: Effects of repeated administration of a combined physiological and psychological stressor," *Psychoneuroendocrinology*, vol. 45, pp. 119–127, Jul. 2014, doi: 10.1016/j.psyneuen.2014.03.022.

[13] G. E. Giles, C. R. Mahoney, T. T. Brunyé, H. A. Taylor, and R. B. Kanarek, "Stress effects on mood, HPA axis, and autonomic response: Comparison of three psychosocial stress paradigms," *PLoS ONE*, vol. 9, no. 12, Dec. 2014, Art. no. e113618.

[14] L. Schwabe and H. Schächinger, "Ten years of research with the socially evaluated cold pressor test: Data from the past and guidelines for the future," *Psychoneuroendocrinology*, vol. 92, pp. 155–161, Jun. 2018.

[15] D. Watson, L. A. Clark, and A. Tellegen, "Development and validation of brief measures of positive and negative affect: The PANAS scales," *J. Personality Social Psychol.*, vol. 54, no. 6, pp. 1063–1070, 1988, doi: 10.1037/0022-3514.54.6.1063.

[16] M. Gjoreski, M. Luätrek, M. Gams, and H. Gjoreski, "Monitoring stress with a wrist device using context," *J. Biomed. Informat.*, vol. 73, pp. 159–170, Sep. 2017, doi: 10.1016/j.jbi.2017.08.006.

[17] Y. Maki, Y. Monno, K. Yoshizaki, M. Tanaka, and M. Okutomi, "Inter-beat interval estimation from facial video based on reliability of BVP signals," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 6525–6528.

[18] M. P. Tarvainen, J.-P. Niskanen, J. A. Lipponen, P. O. Ranta-aho, and P. A. Karjalainen, "Kubios HRV–heart rate variability analysis software," *Comput. Methods Programs Biomed.*, vol. 113, no. 1, pp. 210–220, Jan. 2014, doi: 10.1016/j.cmpb.2013.07.024.

[19] J. A. Lipponen and M. P. Tarvainen, "A robust algorithm for heart rate variability time series artefact correction using novel beat classification," *J. Med. Eng. Technol.*, vol. 43, no. 3, pp. 173–181, Apr. 2019, doi: 10.1080/03091902.2019.1640306.

[20] A. J. A. AlBdairi, Z. Xiao, and M. Alghaili, "Identifying ethnics of people through face recognition: A deep CNN approach," *Sci. Program.*, vol. 2020, pp. 1–7, Jul. 2020.

[21] S. Lawrence, C. L. Giles, A. Chung Tsoi, and A. D. Back, "Face recognition: A convolutional neural-network approach," *IEEE Trans. Neural Netw.*, vol. 8, no. 1, pp. 98–113, Jan. 1997, doi: 10.1109/72.554195.

[22] Y.-X. Yang, C. Wen, K. Xie, F.-Q. Wen, G.-Q. Sheng, and X.-G. Tang, "Face recognition using the SR-CNN model," *Sensors*, vol. 18, no. 12, p. 4237, Dec. 2018, doi: 10.3390/s18124237.

[23] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017, doi: 10.1109/TPAMI.2016.2577031.

[24] A. M. Alayba, V. Palade, M. England, and R. Iqbal, "A combined CNN and LSTM model for Arabic sentiment analysis," in *Proc. Int. Cross-Domain Conf. Mach. Learn. Knowl. Extraction*, vol. 11015, Aug. 2018, pp. 179–191.

[25] S. Liao, J. Wang, R. Yu, K. Sato, and Z. Cheng, "CNN for situations understanding based on sentiment analysis of Twitter data," *Procedia Comput. Sci.*, vol. 111, pp. 376–381, 2017, doi: 10.1016/j.procs.2017.06.037.

[26] B. Kayalibay, G. Jensen, and P. van der Smagt, "CNN-based segmentation of medical imaging data," 2017, *arXiv:1701.03056*. [Online]. Available: http://arxiv.org/abs/1701.03056

[27] *GridSearchCV—Scikit-Learn 0.23.2 Documentation, Scikit-Learn*. Accessed: Dec. 21, 2020. [Online]. Available: https://scikit-learn.org/

[28] H. Han, W. Y. Wang, and B. H. Mao, "Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning," in *Advances in Intelligent Computing* (Lecture Notes in Computer Science), vol. 3644, no. 5. Berlin, Germany: Springer, 2005, pp. 878–887.

[29] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Mach. Learn.*, vol. 63, no. 1, pp. 3–42, Apr. 2006, doi: 10.1007/s10994-006-6226-1.

[30] P. Kalra and V. Sharma, "Mental stress assessment using PPG signal a deep neural network approach," *IETE J. Res.*, vol. 4, pp. 1–7, Nov. 2020.

[31] K. Nkurikiyeyezu, A. Yokokubo, and G. Lopez, "The effect of person-specific biometrics in improving generic stress predictive models," 2019, *arXiv:1910.01770*. [Online]. Available: http://arxiv.org/abs/1910.01770

[32] M. Zubair and C. Yoon, "Mental stress assessment via Ultra-Short-Term recordings of photoplethysmographic sensor," in *Proc. 4th Int. Conf. Comput., Commun. Secur. (ICCCS)*, Oct. 2019, pp. 1–7.

[33] K. Awasthi, P. Nanda, and K. V. Suma, "Performance analysis of machine learning techniques for classification of stress levels using PPG signals," in *Proc. IEEE Int. Conf. Electron., Comput. Commun. Technol. (CONECCT)*, Jul. 2020, pp. 1–8.

[34] J. Aigrain, S. Dubuisson, M. Detyniecki, and M. Chetouani, "Person-specific behavioural features for automatic stress detection," in *Proc. 11th IEEE Int. Conf. Workshops Autom. Face Gesture Recognit. (FG)*, May 2015, pp. 1–6.

[35] Y. Shi *et al.*, "Personalized stress detection from physiological measurements," in *Proc. 2nd Int. Symp. Qual. Life Technol.*, 2010.

SAMI ELZEINY (Student Member, IEEE) received the M.Sc. degree in computer science and engineering from Qatar University, in 2013, and the Ph.D. degree in computer science and engineering from Hamad Bin Khalifa University, Doha, Qatar. He has worked in the banking sector for approximately 21 years and has experience in different IT-related fields, such as IT project management, quality control, system analysis, web application design and development, networking, and system administration. His research interests include image processing, signal processing, artificial intelligence, natural language processing, and the semantic web.

MARWA QARAQE (Member, IEEE) received the bachelor's degree in electrical engineering from Texas A&M University at Qatar, Qatar, in 2010, and the M.Sc. and Ph.D. degrees from Texas A&M University at College Station, College Station, TX, USA, in August 2012 and May 2016, respectively. In September 2016, she joined HBKU as a Faculty Member. Her research interests include wireless communications, signal processing, and machine learning. In particular, she has intensive research expertise in digital signal processing of physiological data and is interested in predictive analytics within the health domain. She serves as a Reviewer for several IEEE and technical international journals and conferences. Throughout her academic career, she received several awards.