IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received June 26, 2018, accepted August 13, 2018, date of publication September 3, 2018, date of current version September 28, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2868361

Long Short Term Memory Hyperparameter Optimization for a Neural Network Based Emotion Recognition Framework

BAHAREH NAKISA[1,2], MOHAMMAD NAIM RASTGOO[1,2], ANDRY RAKOTONIRAINY[2], FREDERIC MAIRE[1], AND VINOD CHANDRAN[1], (Senior Member, IEEE)
[1]School of Electrical Engineering and Computer Science, Queensland University of Technology, Brisbane, Qld 4000, Australia
[2]Centre for Accident Research and Road Safety—Queensland, Queensland University of Technology, Brisbane, Qld 4000, Australia

Corresponding author: Bahareh Nakisa (bahareh.nakisa@hdr.qut.edu.au)

This work is supported by the QUT International Postgraduate Research Scholarship (IPRS).

ABSTRACT Recently, emotion recognition using low-cost wearable sensors based on electroencephalogram and blood volume pulse has received much attention. Long short-term memory (LSTM) networks, a special type of recurrent neural networks, have been applied successfully to emotion classification. However, the performance of these sequence classifiers depends heavily on their hyperparameter values, and it is important to adopt an efficient method to ensure the optimal values. To address this problem, we propose a new framework to automatically optimize LSTM hyperparameters using differential evolution (DE). This is the first systematic study of hyperparameter optimization in the context of emotion classification. In this paper, we evaluate and compare the proposed framework with other state-of-the-art hyperparameter optimization methods (particle swarm optimization, simulated annealing, random search, and tree of Parzen estimators) using a new dataset collected from wearable sensors. Experimental results demonstrate that optimizing LSTM hyperparameters significantly improve the recognition rate of four-quadrant dimensional emotions with a 14% increase in accuracy. The best model based on the optimized LSTM classifier using the DE algorithm achieved 77.68% accuracy. The results also showed that evolutionary computation algorithms, particularly DE, are competitive for ensuring optimized LSTM hyperparameter values. Although DE algorithm is computationally expensive, it is less complex and offers higher diversity in finding optimal solutions.

INDEX TERMS Differential evolution, emotion recognition, hyperparameter optimization, long short term memory, wearable physiological sensors.

I. INTRODUCTION

Automatic emotion recognition using miniaturized physiological sensors and advanced mobile computing technologies has become an important field of research in Human Computer Interaction (HCI) applications. Wearable technologies, such as wireless headbands and smart wristbands, record different physiological signals in an unobtrusive and non-invasive manner. The recorded physiological signals such as Electroencephalography (EEG), Blood Volume Pulse (BVP) and Galvanic Skin Response (GSR) are able to continuously record internal emotional changes. The application of these technologies can be found in detecting driving drowsiness [1], and more recently in assessing the cognitive load of office workers in a controlled environment [2].

However, building a reliable emotion classification system to accurately classify different emotions using physiological sensors is a challenging problem. This is due to the fact that physiological signals are characterized by non-stationarities and nonlinearities. In fact physiological signals consist of time-series data with variation over a long period of time and dependencies within shorter periods. To capture the inherent temporal structure within the physiological data and recognize emotion signature which is reflected in short period of time, we need to apply a classifier which considers temporal information [3], [4].

In recent years, the application of Recurrent Neural Networks (RNNs) for human emotion recognition has led to a significant improvement in recognition accuracy by modeling temporal data. RNNs algorithms are able to elicit the context of observations within sequences and accurately classify sequences that have strong temporal correlation.

However, RNNs have limitations in learning time-series data that stymied their training. Long Short Term Memory networks (LSTM) are a special type of RNNs that have the capability of learning longer temporal sequences [5]. For this reason LSTM networks offer better emotion classification accuracy over other methods when using time-series data [4], [6]–[8].

Although the performance of LSTM networks in classifying different problems is promising, training these networks like other neural networks [9], [10] depend heavily on a set of hyperparameters that determine many aspects of algorithm behavior. It should be noted that there is no generic optimal configuration for all problem domains. Hence, to achieve a successful performance for each problem domain such as emotion classification, it is essential to optimize LSTM hyperparameters. The hyperparameters range from optimization hyperparameters such as number of hidden neurons and batch size, to regularization hyperparameters (weight optimization).

There are two main approaches for hyperparameter optimization: manual and automatic. Manual hyperparameter optimization has relied on experts, which is time consuming. In this approach, the expert interprets how the hyperparameters affect the performance of the model. Automatic hyperparameter optimization methods removes expert input but are difficult to apply due to their high computational cost. Automatic algorithmic approaches range from simple Grid search and Random search to more sophisticated model-based approaches. Grid search, which is a traditional hyperparameter optimization method, is an exhaustive search. Based on this approach, Grid search explores all possible combination of hyperparameters values to find the global optima. Random search algorithms, which are based on direct search methods, are easy to implement. However, these algorithms are converged slowly and take a long time to find the global optima.

Another type of random search methods with ability to converge faster and find an optimal/ near optimal solution in an acceptable time is Evolutionary computation algorithms (EC). EC algorithms such as Particle Swarm Optimization (PSO), and Simulated Annealing (SA) have been shown to be very efficient in solving challenging optimization problems [11], [12]. Among ECs, Differential Evolution (DE) has been successful in different domains due to its capability of maintaining high diversity in exploring and finding better solutions compared to other ECs [13], [14].

In this study, we introduce the use of DE algorithm to optimize LSTM hyperparameters and demonstrate its effectiveness in tuning LSTM hyperparameters to build an accurate emotion classification model. This study focuses on optimizing the number of hidden neurons and batch size for LSTM classifier.

The performances of the optimized LSTM classifiers using the DE algorithm and other state-of-the-art hyperparameter optimization techniques are evaluated on a new dataset collected from wearable sensors. In the new dataset, a lightweight wireless EEG headset (Emotiv) and smart wristband (Empatica E4) are used which meet the consumer criteria for wearability, price, portability and ease-of-use. These new technologies enable the application of emotion recognition in multiple areas such as entertainment, e-learning and the virtual world.

To understand different emotions, there are two existing models of emotions: categorized model and dimensional model. Categorized model is based on a limited number of basic emotions which can be distinguished universally [15]. On the other hand, dimensional emotion divides the emotional space into two to three dimensions, arousal, valence and dominance [16]–[18]. In this study, we categorize different emotional states based on arousal and valence into four quadrants: 1-High Arousal-Positive emotions (HA-P); 2-Low Arousal-Positive emotions (LA-P); 3-High Arousal-Negative emotions (HA-N); 4-Low Arousal-Negative emotions (LA-N).

In summary, the contribution of this study is:

- A new emotion classification framework based on LSTM hyperparameters optimization using the DE algorithm. This study aims to show that optimizing LSTM hyperparameters (batch size and number of hidden neurons) using DE algorithm and selecting a good LSTM network can result in accurate emotion classification system.
- Performance evaluation and comparison of the proposed model with state-of-the-art hyperparameter optimization methods (PSO, SA, Random search and TPE) using a new dataset collected from wireless wearable sensors (Emotiv and Empatica E4). We show that our proposed method surpass them on four-quadrant dimensional emotion classification.

The rest of this paper is structured as follows: Section II presents an overview of related work. The methodology is presented in Section III and experimental results in Section IV. Finally, Section V discusses conclusion and suggestions for future work.

II. RELATED WORK

Emotions play a vital role in our everyday life. Over the past few decades, research has shown that human emotions can be monitored through physiological signals like EEG, BVP and GSR [19] and physical data such as facial expression [20]. Physiological signals offer several advantages over physical data due to their sensitivity for inner feelings and insusceptibility to social masking of emotions [21]. To understand inner human emotions, emotion recognition methods focus on changes in the two major components of nervous system; the Central Nervous System (CNS) and the Autonomic Nervous system (ANS). The physiological signals originating from these two systems carry information relating to inner emotional states.

Gathering physiological signals can be done using two types of sensors: tethered-laboratory and wireless physiological sensors. Although tethered-laboratory sensors are effective and take strong signals with higher resolution, they are more invasive and obtrusive and cannot be used in everyday situations. Wireless physiological sensors can provide a non-invasive and non-obtrusive way to collect physiological signals and can be utilized while carrying out daily activities. However, the resolution of collected signals from these sensors are lower than tethered-laboratory ones.

Among the physiological signals, EEG and BVP signals have been widely used to recognize different emotions, with evidence indicating a strong correlation between emotions such as sadness, anger, surprise and these signals [22], [23]. EEG signals, which measure the electrical activity of the brain, can be recorded by electrodes placed on the scalp. The strong correlation between EEG signals and different emotions is due to the fact that these signals come directly from the CNS, capturing features about internal emotional states. Several studies have focused on extracting features from EEG signals, with features such as time, frequency and time-frequency domains used to recognize emotions. In [14], we proposed a comprehensive set of extractable features from EEG signals, and applied different evolutionary computation algorithms to feature selection to find the optimal subset of features and channels. The proposed feature selection methods are compared over two public datasets (DEAP and MAHNOB) [24], [25], and a new collected dataset. The performance of emotion classification methods are evaluated on DEAP and MAHNOB datasets, using EEG sensors with 32 channels, and compared with the new collected dataset, used a lightweight EEG device with 5 channels. The result showed that evolutionary computation algorithms can effectively support feature selection to identify the salient EEG features and improve the performance of emotion classification. Moreover, the combination of time domain and frequency domain features improved the performance of emotion recognition significantly compared to using only time, frequency and time-frequency domains features. The feasibility of using the lightweight and wearable EEG headbands with 5 channels for emotion recognition is also demonstrated. The use of everyday technology such as lightweight EEG headbands has also proved to be successful in other non-critical domains such as game experience [26], motor imagery [27], and hand movement [28].

Another physiological activity which correlates to different emotions is Blood Volume Pulse (BVP). BVP is a measure that determines the changes of blood volume in vessels and is regulated by ANS. In general, the activity of ANS is involuntarily modulated by external stimuli and emotional states. There are some studies that have investigated BVP features in emotional states and the correlation between them [22], [29], [30]. BVP is measured by Photoplethysmography (PPG) sensor, which senses changes in light absorption density of skin and tissue when illuminated. PPG is a non-invasive and low-cost technique which has recently been embedded in smart wristbands. The usefulness of these wearable sensors has been proven in applications such as stress prediction [31] as well as emotion recognition [22].

Research on wearable physiological sensors to assess emotions has focused on building an accurate classifier based on the advanced machine learning techniques.

Recently, advanced machine learning techniques have achieved empirical success in different applications such as neural machine translation system [32], stress recognition using breathing patterns [33] and emotion recognition using audio-visual inputs [34], [35]. Among these machine learning techniques, RNNs as dynamic models, have achieved state-of-the-art performance in many technical applications [36]–[40], due to their capability in learning sequence modeling tasks.

An RNN is a neural network with cyclic connections, with the ability to learn temporal sequential data. These internal feedback loops in each hidden layer allow RNN networks to capture dynamic temporal patterns and store information. A hidden layer in an RNN contains multiple nodes, which generate the outputs based on the current inputs and the previous hidden states. However, training RNNs is challenging due to vanishing and exploding gradient problems which may hinder the network's ability to back propagate gradients through long-term temporal intervals. This limits the range of contexts they can access, which is of critical importance to sequence data. To overcome the gradient vanishing and exploding problem in RNNs training, LSTM networks were introduced [5]. These networks have achieved top performance in emotion recognition using multi-modal information [3], [41]–[43].

The LSTM cells contain a memory block and gates that let the information go through the connection of the LSTM. There are several connections into and out of these gates. Each gate has its own parameters that need to be trained. In addition to these connections, there are other hyperparameters such as the number of hidden neurons and batch size that need to be selected. Achieving good or even state-of-the-art performance with LSTM requires the selection and optimization of the hyperparameters, and there is no generic optimal configuration for all problem domains. It also requires a certain amount of practical experience to set the hyperparameters. Hence, LSTM can benefit from automatic hyperparameters optimization to help improve the performance of the architecture. This study focuses on optimizing the number of hidden neurons and batch size for LSTM networks.

Hyperparameter optimization can be interpreted as an optimization problem where the objective is to find a value that maximizes the performance and yields a desired model. Sequential model-based optimization (SMBO) is one of the current approaches that is used for hyperparameter optimization [44], [45]. One of the traditional approaches for hyperparameter optimization is grid search. This algorithm is based on an exhaustive searching. However, this algorithm takes a long period of time find the global optima. It has been shown that Random search can perform better than Grid search in optimizing hyperparameters [46]. Strategies such as Tree-based optimization methods, sequentially learn the hyperparameter response function to find the promising next hyperparameter combination.

Recently, several libraries for hyperparameter optimization have been introduced. One of the libraries which provides different algorithms for hyperparameter optimization for machine learning algorithms is the Hyperopt library [47]. Hyperopt can be used for any SMBO problem, and can provide an optimization interface that distinguishes between a configuration space and an evaluation function. Currently, a three algorithms are provided: Random search, Tree-of-Parzen-Estimators (TPE) and Simulated Annealing (SA).

Random search, which is based on direct search methods, is popular because it can provide good predictions in low-dimensional numerical input spaces. Random search first initializes random solutions and then computes the performance of the initialized random solutions using a fitness function. It computes new solutions based on a set of random numbers and other factors, and evaluates the performance of the new solutions using the specified fitness function. Finally, it selects the best solution based on the problem objective (minimization or maximization). Although this algorithm is easy to implement, it suffers from slow convergence.

TPE, which is a non-standard bayesian-based optimization algorithm, uses the tree-structure Parzen estimators for modeling accuracy or error estimation. This algorithm is based on non-parametric approach. Tree-based approaches are particularly suited for high-dimensional and partially categorical input spaces and construct a density estimate over good and bad solutions of each hyperparameter. In fact, TPE algorithm divides the generated solutions into two groups. The first group contains the solutions that improve the current solutions, while the second group contains all other solutions. TPE tries to find a set of solutions which are more likely to be in the first group.

SA method is another useful optimization method, which is inspired by the physical cooling process. There is a gradual cooling process in SA algorithm. This algorithm starts with random solutions (high temperature). This method iteratively generates neighbor solutions given the current solutions and accept the solutions with higher accuracy (lower energy). In the early iteration, the diversity of algorithm in generating new solutions is high. As the number of iterations increases the temperature decreases and the diversity of new solutions decreases. This process makes the algorithm effective in finding the optimum solutions.

Besides SMBO approaches, there are also existing strategies to optimize hyperparameters based on ECs [30]–[32]. EC algorithms such as Particle Swarm Optimization (PSO), and Differential Evolution (DE) are beneficial because they are conceptually simple and can often achieve highly competitive performance in different domains [11], [49], [50]. The advantage of using these algorithms is that while they are generating different solutions with high diversity, they can converge fast. Generating different solutions with high diversity enable the algorithms to find the potential regions with the possibility of optimum solutions. And the fast convergence property makes the algorithms to focus on the potential regions and find the optimal/ near optimal solutions in an acceptable time. It has been shown that PSO is efficient in finding the optimal number of input, hidden nodes and learning rate on time series prediction problems [48]. Although some studies have applied EC algorithms for hyperparameter optimization problems, especially on deep learning algorithms, this is the first study that utilizes DE to optimize LSTM hyperparameters on emotion recognition.

III. METHODOLOGY

In this section, the proposed approach to optimize LSTM hyperparameters using DE algorithm is presented. To optimize LSTM hyperparameters for emotion classification, three main tasks, pre-processing, feature extraction, optimizing hyperparameters of LSTM classifier are considered (see Fig. 1). In this study, we used the most effective pre-processing techniques as well as an effective set of features to keep the processes before classification to the best practice based on previous studies. We then focused on optimizing the performance of the LSTM with optimally selected hyperparameters. The proposed approach is evaluated on our dataset collected using wearable physiological sensors, which is described in the following section. The expected advantages of these sensors include their light-weight, and wireless nature, and are considered highly suitable for naturalistic research.

A. DESCRIPTION OF THE NEW DATASET

The dataset used in this study was collected from 20 participants, aged between 20 and 38 years, while they watched a series of video clips. To collect EEG and BVP signals, Emotiv Insight wireless headset and Empatica E4 were used respectively (see Fig. 2). This Emotiv headset contains 5 channels (AF3, AF4, T7, T8, and Pz) and 2 reference channels located and labeled according to the international 10-20 system (see Fig. 3). TestBench software and Empatica Connect were used for acquiring raw EEG and BVP signals from the Emotiv Insight headset and Empatica respectively.

Emotions were induced by video clips, used in the MAHNOB dataset, and the participants' brain and heart responses were collected while they were watching 9 video clips in succession. The participants were asked to report their emotional state after watching each video, using a keyword such as neutral, anxiety, amusement, sadness, joy or happiness, disgust, anger, surprise, and fear. Before the first video clip, the participants were asked to relax and close their eyes for one minute to allow their baseline EEG and BVP to be determined. Between each video clip stimulus, one minute's silence was given to prevent interference from the previous emotion. The experimental protocol is shown in Fig. 4.

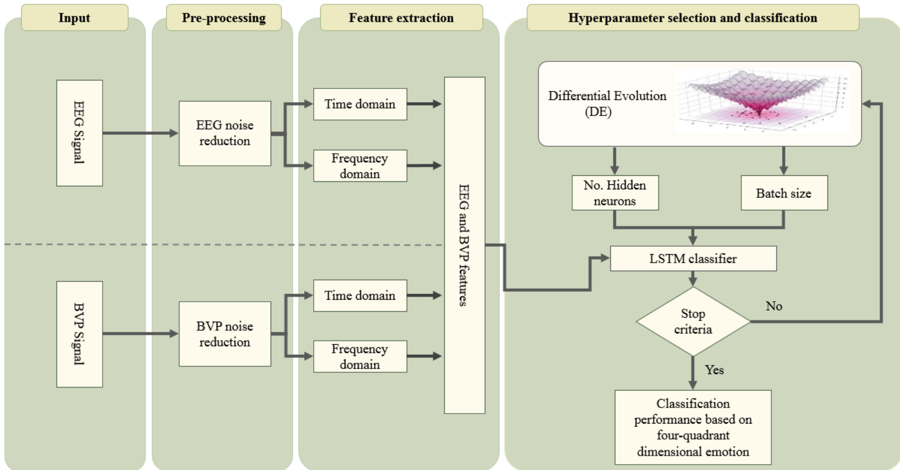


FIGURE 1. The framework to optimize LSTM hyperparameters using DE algorithm for emotion classification.

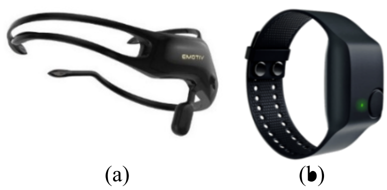


FIGURE 2. (a) The Emotiv Insight headset, (b) the Empatica wristband.

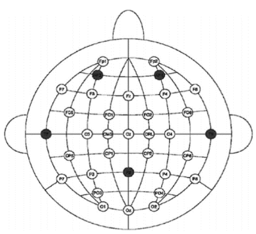


FIGURE 3. The location of five channels in used in emotive sensor (represented by black dot).

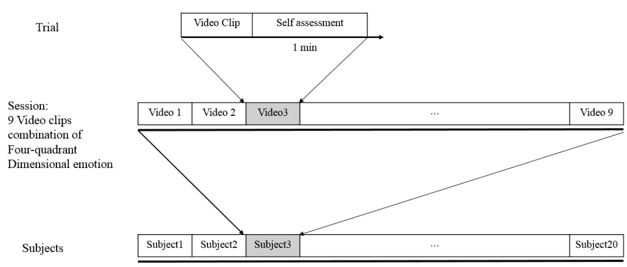


FIGURE 4. Illustration of the experimental protocol for emotion elicitations with 20 participants. Each participant watched 9 video clips and were asked to report their emotions (self-assessment).

To ensure data quality, the signal quality for each participant was manually analyzed. Some EEG signals from the 5 channels were either lost or found to be too noisy due to the long study duration, which may have been caused by loose contact or shifting electrodes. As a result, signal data from 17 (9 female and 8 male) out of 20 participants were included in the dataset. Despite data issues, the experiment allows an investigation into the feasibility of using Emotiv Insight and Empatica E4 for emotion classification purposes.

B. PRE-PROCESSING

To build the emotion recognition model, the critical first step is pre-processing, as EEG and BVP signals are typically contaminated by physiological artifacts caused by electrode movement, eye movement or muscle activities, and heartbeat.

The artifacts generated from eye movement, heartbeat, head movement and respiration are below the frequency rate of 4Hz, while those caused by muscle movement are higher than 40Hz. In addition, there are non-physiological artifacts caused by power lines with frequencies of 50Hz, which can also contaminate the EEG signals. In order to remove artifacts while keeping the EEG signals within specific frequency bands, sixth-order (band-pass) Butterworth filtering was applied to obtain 4-64Hz EEG signals and cover different emotion-related frequency bands. Notch filtering was applied to remove 50Hz noise caused by power lines. In addition to these pre-processing methods, independent component analysis (ICA) was used to reduce the artifacts caused by heartbeat and to separate complex multichannel data into independent components [51], and provide purer signals for feature extraction. To remove noise from the BVP signal, a 3 Hz low pass-Butterworth filter was applied. Fig. 5, 6 show the EEG and BVP signals before and after pre-processing task.

C. FEATURE EXTRACTION

1) EEG FEATURE EXTRACTION

In [14], the combination of time domain and frequency domain features showed more efficient performance compared to only time domain, frequency domain and

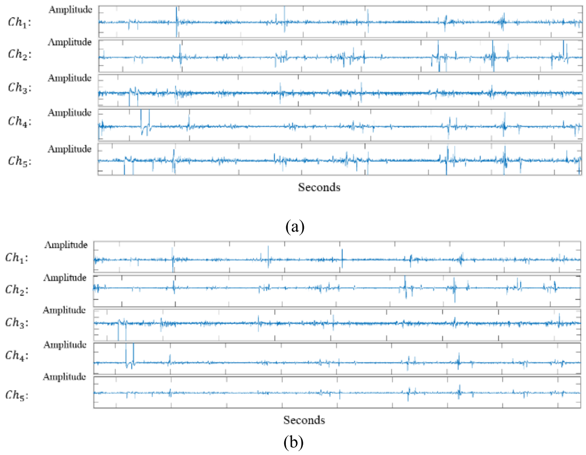


FIGURE 5. **(a) Raw EEG signals before pre-processing, (b) EEG signals after pre-processing.**

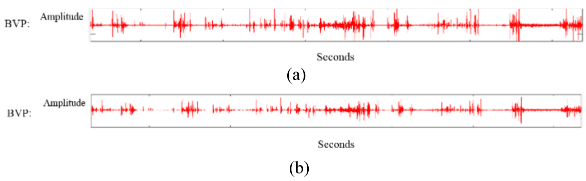


FIGURE 6. **(a) Raw BVP signal before pre-processing, (b) BVP signal after pre-processing.**

time-frequency domain features. Therefore, in order to optimize the process of the feature extraction in this study, we have extracted features from time domain and frequency domain of EEG signals. In addition, one-second window size with 50% overlap is considered for feature extraction. It has been shown that this window size performs well on and is sufficient for capturing emotion recognition [3], [52].

Time domain features have been shown to correlate with different emotional states. Statistical features – such as mean, maximum and minimum values, power, standard deviation, 1st difference, normalized 1st difference, standard deviation of 1st difference, 2nd difference, standard deviation of 2nd difference, normalized 2nd difference, quartile 1, median, quartile 3, quartile 4 can help to classify basic emotions such as joy, fear, sadness [53], [54]. Other promising time domain features are Hjorth parameters: Activity, Mobility and Complexity [55], [56]. These parameters represent the mean power, mean frequency and the number of standard slopes from the signals and have been used in EEG-based studies on sleep disorder and motor imagery [57]–[59]. These features have been applied to real-time applications, as they have the least complexity compared with other methods [60]. In addition to these well-known features, fractal dimension (Higuchi method), which represents the geometric complexity are employed in this study [16], [61], [62]. Non-stationary Index (NSI) segments EEG signals into smaller parts and estimates the variation of their local average to capture the degree of the signals' non-stationarity [63]. Compared to time domain features, frequency domain features have been shown to be more effective for automatic EEG-based emotion recognition. The power of EEG signals among different frequency bands is a good indicator of different emotional states [3]. Features such as power spectrum are extracted from different frequency bands, namely Gamma (30-64 Hz), Theta (13-30 Hz), Alpha (8-13 Hz) and Beta (4-8 Hz), as these features have been shown to change during different emotional states [24], [64]–[66].

2) BLOOD VOLUME PULSE (BVP) FEATURE EXTRACTION

In this study, we employed the BVP signal from the PPG sensor of the Empatica E4 wristband. This sensor measures the relative blood flow in the hands with near infrared light, using photoplethysmography. From the raw BVP signal, we calculated the power spectrum density from three sub-frequency from the range of VLF (0-0.04Hz), LF (0.05-0.15Hz) and HF (0.16-0.4Hz), and the ratio of LF/HF [67]. In addition, typical statistics features such as mean, standard deviation, and variance are calculated from time domain.

D. OPTIMIZING LSTM HYPERPARAMETERS

The main objective is to optimize the hyperparameters of the LSTM classifier using the DE algorithm and achieve better performance on emotion classification. In this study, the parameters in the framework are: batch size and the number of hidden neurons. The DE algorithm starts from the initial solutions (initializes the hyperparameters), which is randomly generated, and then attempts to improve the accuracy of emotion classification model iteratively until stopping criteria is met. The fitness function is LSTM networks which are responsible for performing the evaluation and returning the accuracy of emotion classification.

E. DIFFERENTIAL EVOLUTION (DE)

Proposed by Storn [68], DE was developed to optimize real parameters and real value functions. This algorithm, which is a population-based search, is widely used for continuous search problems [69]. Recently, its strength has been shown in different applications such as strategy adaptation for global numerical optimization [70] as well as feature selection for emotion recognition [14]. Like Genetic Algorithm, DE also uses the crossover and mutation concepts, but it has explicit updating equation. The optimization process in DE consists of four steps: *initialization, mutation, cross over* and *selection.*

In the first step, *initialization,* the initial population $S_i^t = \{s_{1,i}^t, s_{2,i}^t, \ldots, s_{D,i}^t\}$, $i = 1, \ldots, Np$ is randomly generated, where Np is the population size and t shows the current iteration, which in the *initialization* step is equal to zero. It should be mentioned that for each dimension of the problem space there may be some integer ranges that are constrained by some upper and lower bounds: $s_j^{low} \leq s_{j,i} \leq s_j^{up}$, for $j = 1, 2, \ldots D$, where D is the dimension of the problem space. As this study focuses on optimizing the number of hidden neurons and batch size for LSTM networks, therefore, the dimension of the problem is two ($D = 2$).

Each vector forms a candidate solution s_i^t, called target vector, to the multidimensional optimization problem. In the second step, *mutation step*, three individual vectors $s_{j,p}^t$, $s_{j,r}^t$, $s_{j,q}^t$ from population S_i^t are selected randomly to generate a new donor vector v using the following mutation equation:

$$\mathrm{v}_{j,i}^t = s_{j,p}^t + F_i * (s_{j,r}^t - s_{j,q}^t) \tag{1}$$

Where F_i is a constant from [0, 2] denotes the mutation factor. It should be noted that the value of $\mathrm{v}_{j,i}^t$ for each dimension is rounded to the nearest integer value.

In the third step, *crossover step*, the trial vector is computed from each of D dimension of target vector s_i^t and each of D dimension of donor vector v_i^t using the following binomial crossover. In this step, every dimension of the trial vector is controlled by c_r, *crossover rate*, which is a user specified constant within the range [0, 1).

$$\mathrm{u}_{j,i}^t = \begin{cases} v_{i,j}^t & if\ r_i \le c_r\ or\ j = J_{rand} \\ s_{i,j}^t & otherwise \end{cases} \tag{2}$$

Where J_{rand} is a uniformly distributed random integer number in range [1, D], and r_i is a distributed random number in range [0, 1]. In the final step, *selection step*, the target vector s_i is compared with trial vector u_i, and the one with higher accuracy value or lower loss function is selected and admitted to the next generation. It should be noted that in this study the process of selection is achieved by LSTM algorithm, and the vector with the better fitness value is kept. The last three steps continue until specified stopping criteria (300 iterations or reaching 100% accuracy) is met. The following Pseudo-code shows the steps of the DE algorithm as a hyperparameter optimization method (Fig. 7).

F. CONVENTIONAL LSTM

In this work, we applied a conventional LSTM network with two stacked memory block as a fitness function of the DE algorithm. LSTM network evaluates the performance of the emotion classification system based on the obtained values for the number of hidden neurons and batch size. The memory blocks contain memory cells with self-connections storing the temporal state of the network in addition to special multiplicative units called gates to control the flow of information [5]. Each memory block in the original architecture contains three gates: input gate, forget gate, and output gate. The input gate controls the flow of input activation into the memory cells. The forget gate scales the internal state of the cell before adding it as input to the cell through the self-recurrent connection of the cell, therefore adaptively forgetting or resetting the cell's memory. Finally the output gate controls the output flow of cell activation into the next layer.

IV. EXPERIMENTAL RESULTS

We conducted extensive experiments to determine if the DE algorithm can be used as an effective hyperparameter optimization algorithm to improve the performance of the emotion classification using EEG and BVP signals. The performance of the optimized LSTM by the DE algorithm is

Pseudo-code for the DE algorithm with LSTM classifier

Define the size of the population *NP*, *D* dimension of problem, crossover rate c_r, scale factor *F*.

Initialization: Initialize the population $S_i^{t=0} = \{s_{1,i}^t, s_{2,i}^t, \ldots, s_{D,i}^t\}$, $i = 1, \ldots, Np$ *which each individual uniformly distributed in the range*$[s^{low}, s^{high}]$

While the termination criteria is not met

For each individual, target vector, in the population NP

Mutation: Select three individual from the population randomly and generate a donor vector v_i^t using the following mutation equation:

$\mathrm{v}_{j,i}^t = \mathrm{s}_{j,p}^t + \mathrm{F}_i * (\mathrm{s}_{j,r}^t + \mathrm{s}_{j,q}^t)$

Crossover: Compute the trial vector for the *i*th target vector $\mathrm{u}_{j,i}^{t+1}$:

$\mathrm{u}_{j,i}^t = \begin{cases} v_{i,j}^t & if\ \ r_i \le c_r\ or\ j = J_{rand} \\ s_{i,j}^t & \quad otherwise \end{cases}$

Selection: Apply LSTM classifier as fitness function *f* and evaluate s_i^t and u_i^t:

If $f(s_i^t) \le f(\mathrm{u}_i^t)$ *then* $s_i^{t+1} = \mathrm{u}_i^t$

Else $s_i^{t+1} = s_i^t$

End For

End While

FIGURE 7. The Pseudo-code of the LSTM hyperparameter optimization using DE algorithm.

compared with the other well-known hyperparameter optimization algorithms (PSO, SA, Random search and TPE).

The experiments focused on classifying the four-quadrant dimensional emotions (HA-P, LA-P, HA-N and LA-N) based on the optimized LSTM. The performance of the proposed system is evaluated on the data collected from 17 participants while they were watching 9 video clips. It should be mentioned that the length of each video clip varied and in order to prepare data for LSTM network with variable length, we transferred data with the same length. In this case, we considered the maximum length and applied the zero-padding to pad variable length.

Before applying zero-padding, noise reduction techniques such as Butterworth, notch filtering and ICA were applied. After applying noise reduction and zero-padding, different features (time-domain, frequency-domain) from one second window with a 50% overlap were extracted from 5 EEG channels and a BVP signal. Twenty-five features from each window of each EEG channel and 8 features from each window of BVP signal were extracted.

These features were then concatenated to form a vector (EEG + BVP = 25 × 5 + 8 × 1 = 133 features) and fed into LSTM network. The LSTM network consisted of two stacked cells developed in TensorFlow. The number of input nodes corresponded to the number of extracted features from EEG and BVP signal (133 features) and the number of output corresponded to the number of target classes (4 classes).

The LSTM network used the following settings:

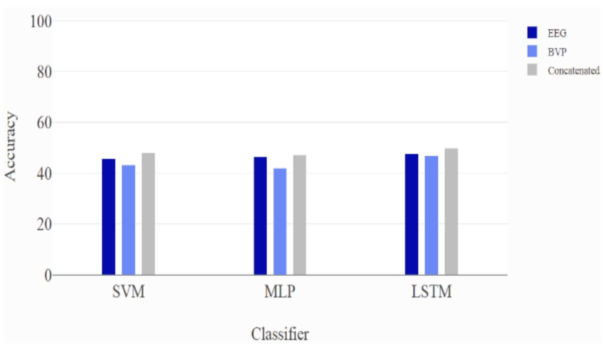
$$\text{lambda\_loss\_amount} = 0.0015, \quad \text{learning rate} = 0.0025,$$
$$\text{training epoch} = 10, \quad \text{number of inputs} = 133.$$

The results is produced using 17 leave-one-subject out cross validation. In this method one participant is used for testing and the remaining participants are used for training. Then the classification model was built for the training dataset and the test dataset was classified using this model to assess the accuracy. This process was repeated 17 times using different participants as test dataset. The 17 leave-one-subject out cross validation trial is repeated 5 times to obtain a statistically stable performance of the recognition system. The overall emotion recognition performance (accuracy/ valid loss) is obtained by averaging the result of 5 times cross validation trials.

A. FUSION OF EEG AND BVP

In this section, we assess the performance of emotion classification using EEG signals, BVP signal and their fusion. To evaluate the performance of the fusion of EEG and BVP signals, we applied feature level fusion.

Fig. 8 presents the performance of each modality as well as the fusion of these two modalities using three well-known classifiers in emotion recognition (SVM, Multilayer Perceptron (MLP) and LSTM). In the previous sections, we chose to have a LSTM network with two stacked cells. The number of hidden neurons in this configuration is set to a quarter of the input layer of the features (133/4 = 33), and the batch size is set to 100.



Classifier	**EEG**	**BVP**	**EEG+BVP**
SVM	**45.5%**	43.1%	47.9%
MLP	**46.2%**	41.9%	47.1%
LSTM	**47.5%**	46.8%	**49.6%**

FIGURE 8. **The performance (classification accuracy) of each signal and their fusion at feature level using different classifiers.**

The result showed that the performance of LSTM classifier based on the fusion of EEG and BVP signals is better compared to SVM and MLP classifiers.

It also showed that EEG signals from Emotiv are performing better than BVP signal generated from Empatica E4, and the performance of fusion of both sensors is higher than the single ones.

B. EVALUATING LSTM HYPERPARAMETERS USING DE AND OTHER METHODS

The performance of the LSTM classifier optimized by the DE algorithm is evaluated in this section. It is also compared with other existing methods (PSO, SA, Random search and TPE) based on the optimum accuracy and valid loss achieved within reasonable time. To apply Random search, TPE and SA, Hyperopt library was used, and the DE algorithm is developed in Python language programming. The following settings are applied on each hyperparameter optimization algorithm:

For the DE algorithm, the crossover probability is set to 0.2, bounds for batch size is between 1 and 271, bounds for the number of hidden neurons are set between 1 and 133 and the population size is equal to 5. For the SA algorithm with Hyperopt library, anneal.suggest algorithm is selected, bounds for batch size are placed between [1- 271] and the number of hidden neurons are set between [1- 133]. For the Random search with Hyperopt library, Random.suggest algorithm is chosen. Batch size is selected from the range of 1 and 271, and the number of hidden neurons are ranges from 1 to 133. For the TPE search with Hyperopt library, tpe.suggest is used, bound for batch size is between [1- 271], and the number of hidden neurons are ranges from 1 to 133. In order to find an optimized LSTM to classify emotions more accurately, we ran each hyperparameter optimization algorithms with three different iterations (50, 100 and 300 iterations). These iterations assisted the algorithms to explore and find more solutions (maximum 1500 solutions $= 5 \times 300$, $5 =$ population size, $300 =$ number of iterations). In this experiment, each hyperparameter optimization algorithm is tested based on its ability to achieve the best values for the number of hidden neurons and batch size for LSTM classifier to improve the performance of the proposed system.

Table 1 presents the performance of all hyperparameter optimization methods, providing processing time, average accuracy $\pm$ standard deviation, best accuracy, average loss value $\pm$ standard deviation, and the best loss found in different iterations. The presented mean accuracy, average time, mean loss value and their standard deviation are the result of averaging 5 times cross validation trials (17 one-leave-subject-out cross validation). Processing time is determined by Intel Core i7 CPU, 16 GB RAM, running windows 7 on 64- bit architecture.

Based on the average time, the processing time for SA algorithm was lower than the others, while its performance was lower than PSO and the DE algorithms and higher than Random search and TPE algorithms. In contrast, the DE took the most processing time over all iterations, however its performance was higher than all other algorithms.

Since the goal of this study is to introduce an optimized LSTM model that can accurately classify different emotions the computational time to build such an optimized classification model is less important, as long as it can achieve an acceptable result. In fact, finding the optimal classifier is done offline, and the time (358 hours for 17 cross fold validation at 300 iterations) to do this is within reasonable development time for such projects and will be considerably less with higher performance, parallel or cloud computing. Based on the best accuracy, LSTM classifier using the DE algorithm

TABLE 1. The average accuracy, time, valid loss and best loss of different hyperparameter optimization methods.

Hyperparameter optimization	No. iteration	Time	Accuracy (mean± std)	Best Accuracy	Valid loss (mean± std)	Best loss
Random search	50	25 h	53.47 ±15.9	67.35	22.47 ± 6.8	1.9
	100	58 h	59.50 ±12.9	68.59	21.68 ± 6.5	**1.1**
	300	174 h	59.21 ±16.1	68.63	21.39 ±11.9	1.5
TPE	50	29 h	58.16 ±12.1	64	24.48 ±5.1	11.3
	100	57 h	59.95 ±12.1	68.5	19.64 ±5.8	5.2
	300	172 h	60.86 ±14.46	68.9	17.45 ±8.9	3.1
SA	50	23 h	50.17±17.4	59.65	21.72±5.5	1.96
	100	47 h	60.98±12.4	68.65	21.05±5.1	1.73
	300	143 h	62.35±12.06	69.1	20.76±9.1	1.91
PSO	50	25 h	61.32 ±10.7	67.65	16.21 ±2.3	5.43
	100	55 h	**63.99±9.4**	69.98	14.84 ±2.6	3.44
	300	167 h	62.61±11.7	69.65	10.97 ±4.9	2.4
DE	50	59 h	**62.77±10.1**	**68.4**	**14.33±1.3**	**1.2**
	100	119 h	63.91±12.9	**73.36**	**12.65±2.8**	**1.14**
	300	358 h	**67.52±10.3**	**77.68**	**10.57±3.5**	**1.16**

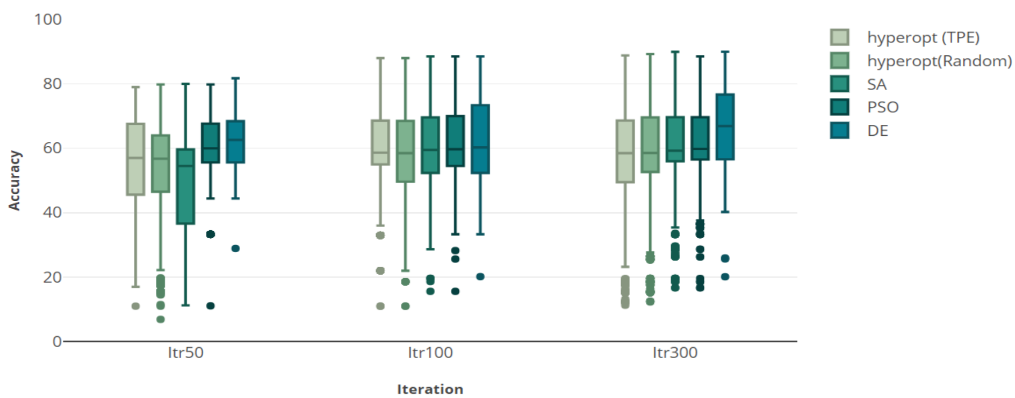


FIGURE 9. Box plots showing the distribution accuracy of the proposed system optmized by DE, PSO, SA, Random search and TPE algorithms in three different configurations on our collected dataset.

achieved significant result (77% accuracy) at 300 iteration, although its computational cost is expensive.

It also demonstrated that the best achieved accuracies of all the other hyperparameter optimization algorithms are less than 70%, and this confirms the ability of the DE algorithm in finding a better solution. Based on the mean accuracy, the performance of DE algorithms followed by PSO and SA algorithms from the early iterations, was significantly better than Random search and TPE algorithms. From the 50 to 100 iterations, the improvement rate for DE and PSO algorithms was 3%, however, after 100 iterations the average accuracy of PSO algorithm decreased while for DE it improved slightly. This phenomenon is most likely due to the DE diversity property and its capabilities in searching and exploring more solutions. However, ECs did not achieve a significant result due to their convergence property and trapping in local optima [71], [72]. Trapping in local optima is a common problem among ECs. Performance differences across different hyperparameter optimization methods were also tested for statistical significance using a two-way repeated measure ANOVA. The results showed that mean accuracy for each iteration (50, 100, 300 iterations) differed significantly between all hyperparameter optimization methods (F-ratio $=$ 36.225, ρ-value $<$ 0.00001).

Further analysis of all algorithms and the relative distribution (accuracy) is shown in Fig. 9. From Fig.9, it is apparent that as the number of iterations increased, the accuracies of LSTM classifier using all hyperparameter optimization methods are improved (the maximum accuracy is improved by about 17%). However, the median accuracy of the DE algorithm over all three iterations is higher than all the other algorithms, particularly in 300 iterations. Results showed that in 50 iterations, the median accuracies of LSTM using DE and PSO algorithms were higher than 60%, while the median accuracies of the other three algorithms (TPE, Random search and SA algorithms) were lower than 60%. Fig. 8 also shows that as the number of iterations increased, no significant improvement using TPE, Random search were found. However, the median accuracy of LSTM using DE was improved by 7%. The result confirmed that

LSTM hyperparameter optimization using DE followed by PSO algorithms is more robust compared to TPE and Random search.

The changing process of algorithms within 300 iterations are presented in Fig. 10. We plot the average test accuracy of each algorithm over 17 cross validation trials. Based on the figure, the performance of all hyperparameters in early iterations are increasing and close to each other. However, after some iterations, the accuracy of TPE, SA and PSO algorithms remain steady with slight improvements. It is shown that as the number of iterations increased, the performance of DE improved more compared to other algorithms. It can find better solutions with higher accuracy, while PSO and SA algorithms stagnated to local optima due to less diversity in the nature of these algorithms.

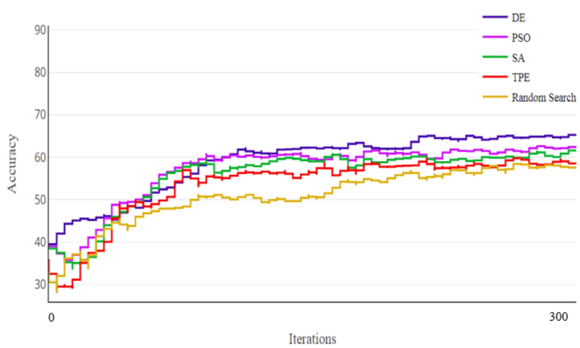


FIGURE 10. The average accuracies of the hyperparameter optimization methods over 300 iterations.

To show that how the optimized LSTM network can classify four-quadrant dimensional emotion and demonstrate the connection between the best achieved performances of the tuned LSTM model and different affects, confusion matrices of such models are presented (Tables 2-7). Table 2 presents the confusion matrix of LSTM without using any hyperparameter optimization algorithm. Using this approach the batch size and the number of hidden neurons are randomly selected. Tables 3-7 show the confusion matrices of the optimized LSTM classifiers using DE, PSO, SA, Random search and TPE algorithms as hyperparameter optimization techniques respectively. The provided confusion matrices show the best tuned LSTM model using the mentioned hyperparameter optimization algorithms at 300 iterations.

TABLE 2. The emotional confusion matrix corresponding to the LSTM models without Hyperparameter optimization.

	HA-P	HA-N	LA-P	LA-N
HA-P	728	363	82	27
HA-N	224	560	74	112
LA-P	186	28	435	271
LA-N	199	95	346	590

TABLE 3. The emotional confusion matrix corresponding to the best achieved LSTM models using DE algorithm.

	HA-P	HA-N	LA-P	LA-N
HA-P	994	153	43	10
HA-N	72	807	23	68
LA-P	98	8	695	119
LA-N	92	28	197	913

TABLE 4. The emotional confusion matrix corresponding to the best achieved LSTM models using PSO algorithm.

	HA-P	HA-N	LA-P	LA-N
HA-P	853	267	64	16
HA-N	136	712	27	95
LA-P	95	15	633	177
LA-N	172	27	236	795

TABLE 5. The emotional confusion matrix corresponding to the best achieved LSTM models using SA algorithm.

	HA-P	HA-N	LA-P	LA-N
HA-P	837	313	33	17
HA-N	123	725	39	83
LA-P	122	13	567	218
LA-N	144	24	212	850

TABLE 6. The emotional confusion matrix corresponding to the best achieved LSTM models using Random search algorithm.

	HA-P	HA-N	LA-P	LA-N
HA-P	860	266	51	23
HA-N	165	657	45	103
LA-P	109	18	596	197
LA-N	186	29	224	791

Based on Table 2, recognizing HA-P, HA-N and LA-N emotions are more difficult using the LSTM algorithm without any hyperparameter optimization algorithm. It is shown that recognizing HA-N is more confused with HA-P and vice versa. LA-P quadrant is mostly misclassified as HA-P and LA-N, and LA-N is misclassified as LA-P and HA-P.

However, the optimized LSTM using the DE algorithm could classify all four-quadrant dimensions better than the other LSTM classifiers without and with hyperparameter optimization algorithm(s). The achieved confusion matrix from the best tuned LSTM classifier using the DE algorithm demonstrate that DE algorithm is able to find better values for the LSTM hyperparameters and improve the performance of

TABLE 7. The emotional confusion matrix corresponding to the best achieved LSTM models using TPE algorithm.

	HA-P	HA-N	LA-P	LA-N
HA-P	811	298	78	13
HA-N	148	675	34	113
LA-P	143	18	582	177
LA-N	166	19	259	786

emotion classification. Based on Table 3, the performance of the LSTM classifier using the DE in recognizing HA-P and HA-N is better than LA-P and LA-N. Although, the LSTM using the DE algorithm could classify these two quadrants better than the other hyperparameter optimization, it is still difficult to classify these two quadrant accurately.

From the other Tables (4-7), we can generally observe that recognizing LA-N using the optimized LSTM classifiers using SA, Random search and TPE is more difficult than the other three quadrants (HA-P, HA-N and LA-P), and this quadrant is mainly misclassified as LA-P and HA-P.

C. COMPARISON WITH OTHER LATEST WORKS

The final experiment compared the performance of our proposed framework (the optimized LSTM by DE) against some other state-of-the-art studies. Experimental results are shown in Table 8, indicating the average classification accuracy for different emotion classification methods based on different number of classes as well as different classifiers.

In addition, we compared the performance of emotion classification methods using tethered-laboratory and wireless physiological sensors. Tethered-laboratory sensors are wired sensors with high quality signals which are usually used for clinical purpose, but they are not deployable for open natural environments.

We used wireless physiological sensors in this study and these types of sensors can be used for real-life situations.

The comparison shows that the average performance of both tethered-laboratory and wireless physiological sensors in two-class emotion classification is significant. However,

TABLE 8. comparison of our approach with some latest works.

Method	Classifier	No. classes	Accuracy	Sensor
Sourina and Y. Liu **[16]**	SVM	2 (Arousal, Valence)	90%	EEG (tethered-laboratory)
Ramirez et.al. **[17]**	SVM(RBF)	2 (Arousal, Valence)	Arousal: 78% Valence: 80%	EEG (Wireless)
G. dovičić et al. **[74]**	SVM, KNN	2 (Arousal, Valence)	Arousal: 67% Valence: 70%	BVP& GSR (tethered-laboratory)
Xu etal. **[75]**	J48,Naïve, KNN, VM	2 (Arousal, Valence)	J48: 60% Naïve: 56% KNN: 52% SVM: 20%	BVP (tethered-laboratory)
Ragot et al. **[74]**	SVM	2 (Arousal, Valence)	Arousal: 65% Valence: 69%	BVP (tethered-laboratory)
			Arousal: 65% Valence: 70%	BVP (Wireless)
Rakshit et al. **[73]**	SVM	3 (Happy, Sad, Neutral)	83%	BVP (tethered-laboratory)
Ackermann et al. **[77]**	SVM, Random forest	3 (Anger, Surprise, other)	55%	EEG (tethered-laboratory)
Feradov and Ganchev **[78]**	SVM	3 (Negative, Positive, Neutral)	62%	EEG (Wireless)
Candra et al. **[18]**	SVM	4 (Sad, Relaxed, Angry, Happy)	60.9±3.2	EEG (tethered-laboratory)
Nakisa, et al. **[14]**	PNN	4 (HP-A, HA-N, LA-P, LA-N	65.04±3.1	EEG (Wireless)
Our work	LSTM	4 (HP-A, HA-N, LA-P, LA-N)	**66.92±9.3** **(Best model: 77.68%)**	EEG & BVP (Wireless)

the performance of EEG signals (tethered-laboratory and wireless) are better than BVP signal. The result also shows that SVM and decision tree can classify emotions into two-class better than the other classifiers.

In three-class emotion classification, Rakshit *et al.* [71] achieved higher accuracy using tethered-laboratory BVP signal, while the performance of EEG signals, tethered-laboratory and wireless, is low. Moreover, Table 8 shows that the obtained accuracies by tethered-laboratory EEG sensor and wireless EEG sensor seem similar. Candra *et al.* [18] classified four different emotions (sad, relaxed, angry and happy) using SVM classifier. The result indicated that the average performance of four-class emotion classification using EEG signals (tethered-laboratory) is promising, which is better than three-class emotion classification.

In [14], we achieved promising accuracy (65% accuracy on average) using the only wireless wearable EEG sensor. It should be stated that the average achieved performance using wireless EEG sensor can compete with tethered-laboratory EEG sensor with higher resolution.

In comparison with other works, our proposed system (optimized LSTM by DE) has shown the-state-of-the-art performance in classifying four-quadrant dimensional emotions using wireless wearable physiological sensors. The average achieved accuracy confirms that fusion of EEG and BVP signals along with the optimized classifier can help in classifying four-class emotions more accurately.

Since the goal of this work is to find an optimized LSTM classifier with high performance, the performance of the best model is also presented in the Table 8. The optimized LSTM classifier using DE algorithm (best model) achieved 77.68% accuracy in classifying four-quadrant dimensional emotions.

V. CONCLUSION

In this study, we presented a new framework based on the use of a DE algorithm to optimize LSTM hyperparameters (batch size and number hidden neurons) in the context of emotion classification. The performance of the proposed framework is evaluated and compared with other state-of-the-art hyperparameter optimization algorithms (PSO, SA, Random search and TPE) over new data collected using lightweight physiological signals (Emotiv and Empatica E4), which can measure EEG and BVP signals. This performance was evaluated based on four-quadrant dimensional emotions: High Arousal Positive emotions (HA-P), Low Arousal-Positive emotions (LA-P), High Arousal-Negative emotions (HA-N), and Low Arousal- Negative emotions (LA-N). The results show that fusion of EEG and BVP signals provided higher performance for classifying four-quadrant dimensional emotions. In addition, the performance of the proposed system using different hyperparameter optimization methods was compared with different time intervals, 50, 100 and 300 iterations. The results demonstrated that that the average accuracy of the system based on the optimized LSTM network using DE and PSO algorithms improved over every time interval. The average accuracy of the proposed framework based on the optimized LSTM network using the DE is compared with other latest works and the result demonstrated that finding good values for LSTM hyperparameters can enhance emotion classification significantly. The better optimized LSTM classifier using the DE algorithm achieved 77.68% accuracy for four-quadrant emotion classification. However, the performance of the best achieved models using the other hyperparameters optimization algorithms is lower than 70% accuracy. This finding confirms the ability of the DE algorithm to search and find better solution compared to the other state-of-the-art hyperparameter optimization methods.

It should be noted that after a number of iterations (100 iterations) the performance of the system using ECs (PSO, SA and DE) did not change significantly. This could be due to the occurrence of premature convergence problem which may cause the swarm to be trapped into local optima and unable to explore other promising solutions. Therefore, it is recommended to explore a new development of EC algorithms to overcome the premature convergence problem and further improve emotion classification performance. In addition, since the processing time of DE is more expensive, this can be reduced using parallel and/or cloud computing.

REFERENCES

[1] G. Li, B.-L. Lee, and W.-Y. Chung, "Smartwatch-based wearable EEG system for driver drowsiness detection," *IEEE Sensors J.*, vol. 15, no. 12, pp. 7169–7180, Dec. 2015.
[2] F. Zhang *et al.*, "The effects of higher temperature setpoints during summer on office workers' cognitive load and thermal comfort," *Building Environ.*, vol. 123, pp. 176–188, Oct. 2017.
[3] M. Soleymani, S. Asghari-Esfeden, Y. Fu, and M. Pantic, "Analysis of EEG signals and facial expressions for continuous emotion detection," *IEEE Trans. Affect. Comput.*, vol. 7, no. 1, pp. 17–28, Jan./Mar. 2016.
[4] M. Wöllmer, M. Kaiser, F. Eyben, B. Schuller, and G. Rigoll, "LSTM-modeling of continuous emotions in an audiovisual affect recognition framework," *Image Vis. Comput.*, vol. 31, no. 2, pp. 153–163, 2013.
[5] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[6] Y. Kim, H. Lee, and E. M. Provost, "Deep learning for robust feature generation in audiovisual emotion recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2013, pp. 3687–3691.
[7] M. Wöllmer *et al.*, "Abandoning emotion classes—Towards continuous emotion recognition with modelling of long-range dependencies," in *Proc. 9th Annu. Conf. Int. Speech Commun. Assoc.*, 2008, pp. 1–4.
[8] F.-S. Tsai, Y.-M. Weng, C.-J. Ng, and C.-C. Lee, "Embedding stacked bottleneck vocal features in a LSTM architecture for automatic pain level classification during emergency triage," in *Proc. 7th Int. Conf. Affect. Comput. Intell. Interact. (ACII)*, Oct. 2017, pp. 313–318.
[9] K. Liu, L. M. Zhang, and Y. W. Sun, "Deep Boltzmann machines aided design based on genetic algorithms," *Appl. Mech. Mater.*, vols. 568–570, pp. 848–851, Jun. 2014.
[10] J. P. Papa, G. H. Rosa, K. A. Costa, N. A. Marana, W. Scheirer, and D. D. Cox, "On the model selection of bernoulli restricted Boltzmann machines through harmony search," in *Proc. Companion Publication Annu. Conf. Genetic Evol. Comput.*, 2015, pp. 1449–1450.
[11] M. N. Rastgoo, B. Nakisa, and M. Z. A. Nazri, "A hybrid of modified PSO and local search on a multi-robot search system," *Int. J. Adv. Robotic Syst.*, vol. 12, no. 7, p. 86, 2015.
[12] S.-M. Chen and C.-Y. Chien, "Solving the traveling salesman problem based on the genetic simulated annealing ant colony system with particle swarm optimization techniques," *Expert Syst. Appl.*, vol. 38, no. 12, pp. 14439–14450, 2011.
[13] M. Z. Baig, N. Aslam, H. P. Shum, and L. Zhang, "Differential evolution algorithm as a tool for optimal feature subset selection in motor imagery EEG," *Expert Syst. Appl.*, vol. 90, pp. 184–195, Dec. 2017.

[14] B. Nakisa, M. N. Rastgoo, D. Tjondronegoro, and V. Chandran, "Evolutionary computation algorithms for feature selection of EEG-based emotion recognition using mobile sensors," *Expert Syst. Appl.*, vol. 93, pp. 143–155, Mar. 2017.

[15] P. Ekman, "Are there basic emotions?" *Psychol. Rev.*, vol. 99, no. 3, pp. 550–553, 1992.

[16] O. Sourina and Y. Liu, "A fractal-based algorithm of emotion recognition from EEG using arousal-valence model," in *Proc. Biosignals*, 2011, pp. 209–214.

[17] R. Ramirez and Z. Vamvakousis, "Detecting emotion from EEG signals using the emotive EPOC device," in *Brain Informatics*, vol. 7670. Berlin, Germany: Springer, 2012, pp. 175–184.

[18] H. Candra, M. Yuwono, A. Handojoseno, R. Chai, S. Su, and H. T. Nguyen, "Recognizing emotions from EEG subbands using wavelet analysis," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 6030–6033.

[19] S. Koelstra *et al.*, "Single trial classification of EEG and peripheral physiological signals for recognition of emotions induced by music videos," in *Proc. Int. Conf. Brain Inform.*, 2010, pp. 89–100.

[20] M. S. Hossain and G. Muhammad, "An emotion recognition system for mobile applications," *IEEE Access*, vol. 5, pp. 2281–2287, 2017.

[21] J. Kim, "Bimodal emotion recognition using speech and physiological changes," in *Robust Speech Recognition and Understanding*, M. Grimm and K. Kroschel, Eds. Vienna, Austria: I-Tech Education, 2007, pp. 265–280.

[22] A. Haag, S. Goronzy, P. Schaich, and J. Williams, "Emotion recognition using bio-sensors: First steps towards an automatic system," in *Proc. Tuts. Res. Workshop Affect. Dialogue Syst.*, Kloster Irsee, Germany: Springer, 2004, pp. 36–48.

[23] K. H. Kim, S. W. Bang, and S. R. Kim, "Emotion recognition system using short-term monitoring of physiological signals," *Med. Biol. Eng. Comput.*, vol. 42, no. 3, pp. 419–427, 2004.

[24] S. Koelstra *et al.*, "DEAP: A database for emotion analysis ;using physiological signals," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18–31, Jan./Mar. 2012.

[25] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 42–55, Jan./Mar. 2012.

[26] T. McMahan, I. Parberry, and T. D. Parsons, "Evaluating electroencephalography engagement indices during video game play," in *Proc. FDG*, 2015, pp. 1–8.

[27] C. Kranczioch, C. Zich, I. Schierholz, and A. Sterr, "Mobile EEG and its potential to promote the theory and application of imagery-based motor rehabilitation," *Int. J. Psychophysiol.*, vol. 91, no. 1, pp. 10–15, 2014.

[28] N. Robinson and A. P. Vinod, "Bi-directional imagined hand movement classification using low cost EEG-based BCI," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Oct. 2015, pp. 3134–3139.

[29] A. M. Khan and M. Lawo, "Recognizing emotion from blood volume pulse and skin conductance sensor using machine learning algorithms," in *Proc. 14th Medit. Conf. Med. Biol. Eng. Comput.*, 2016, pp. 1297–1303.

[30] K. Takahashi, "Remarks on SVM-based emotion recognition from multimodal bio-potential signals," in *Proc. 13th IEEE Int. Workshop Robot Human Interact. Commun.*, Sep. 2004, pp. 95–100.

[31] A. Ghosh, M. Danieli, and G. Riccardi, "Annotation and prediction of stress and workload from physiological and inertial signals," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 1621–1624.

[32] H. Qin, T. Shinozaki, and K. Duh, "Evolution strategy based automatic tuning of neural machine translation systems," in *Proc. 14th Int. Workshop Spoken Language Transl.*, Tokyo, Japan, Dec. 2017, pp. 120–128.

[33] Y. Cho, N. Bianchi-Berthouze, and S. J. Julier, "DeepBreath: Deep learning of breathing patterns for automatic stress recognition using low-cost thermal imaging in unconstrained settings," in *Proc. 7th Int. Conf. Affect. Comput. Intell. Interact. (ACII)*, Oct. 2017, pp. 456–463.

[34] C.-M. Chang and C.-C. Lee, "Fusion of multiple emotion perspectives: Improving affect recognition through integrating cross-lingual emotion information," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 5820–5824.

[35] C.-M. Chang, B.-H. Su, S.-C. Lin, J.-L. Li, and C.-C. Lee, "A bootstrapped multi-view weighted Kernel fusion framework for cross-corpus integration of multimodal emotion recognition," in *Proc. 7th Int. Conf. Affect. Comput. Intell. Interact. (ACII)*, Oct. 2017, pp. 377–382.

[36] A. Graves, A.-R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2013, pp. 6645–6649.

[37] L. Yao *et al.*, "Describing videos by exploiting temporal structure," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 4507–4515.

[38] S. E. Kahou, V. Michalski, K. Konda, R. Memisevic, and C. Pal, "Recurrent neural networks for emotion recognition in video," in *Proc. ACM Int. Conf. Multimodal Interact.*, 2015, pp. 467–474.

[39] A. Graves, *Supervised Sequence Labelling With Recurrent Neural Networks*. Berlin, Germany: Springer-Verlag, 2012.

[40] B. Zoph and Q. V. Le. (2016). "Neural architecture search with reinforcement learning." [Online]. Available: https://arxiv.org/abs/1611.01578

[41] L. Chao, J. Tao, M. Yang, Y. Li, and Z. Wen, "Long short term memory recurrent neural network based multimodal dimensional emotion recognition," in *Proc. 5th Int. Workshop Audio/Vis. Emotion Challenge*, 2015, pp. 65–72.

[42] S. Chen and Q. Jin, "Multi-modal dimensional emotion recognition using recurrent neural networks," in *Proc. 5th Int. Workshop Audio/Vis. Emotion Challenge*, 2015, pp. 49–56.

[43] M. Nicolaou, H. Gunes, and M. Pantic, "Continuous prediction of spontaneous affect from multiple cues and modalities in valence-arousal space," *IEEE Trans. Affect. Comput.*, vol. 2, no. 2, pp. 92–105, Apr. 2011.

[44] J. S. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, "Algorithms for hyper-parameter optimization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, pp. 2546–2554.

[45] F. Hutter, H. H. Hoos, and K. Leyton-Brown, "Sequential model-based optimization for general algorithm configuration," in *Proc. Int. Conf. Learn. Intell. Optim.*, 2011, pp. 507–523.

[46] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *J. Mach. Learn. Res.*, vol. 13, pp. 281–305, Feb. 2012.

[47] J. Bergstra, B. Komer, C. Eliasmith, D. Yamins, and D. D. Cox, "Hyperopt: A Python library for model selection and hyperparameter optimization," *Comput. Sci. Discovery*, vol. 8, no. 1, p. 014008, 2015.

[48] T. Kuremoto, S. Kimura, K. Kobayashi, and M. Obayashi, "Time series forecasting using restricted boltzmann machine," in *Proc. Int. Conf. Intell. Comput.*, 2012, pp. 17–22.

[49] B. Nakisa, M. N. Rastgoo, M. F. Nasrudin, and M. Z. A. Nazri, "A multi-swarm particle swarm optimization with local search on multi-robot search system," *J. Theor. Appl. Inf. Technol.*, vol. 71, no. 1, pp. 129–136, 2014.

[50] B. Nakisa, M. N. Rastgoo, and M. J. Norodin, "Balancing exploration and exploitation in particle swarm optimization on search tasking," *Res. J. Appl. Sci., Eng. Technol.*, vol. 8, no. 12, pp. 1429–1434, 2014.

[51] T. P. Jung *et al.*, "Removing electroencephalographic artifacts by blind source separation," *Psychophysiology*, vol. 37, no. 2, pp. 163–178, Mar. 2000.

[52] D. Le and E. M. Provost, "Emotion recognition from spontaneous speech using Hidden Markov models with deep belief networks," in *Proc. IEEE Workshop Autom. Speech Recognit. Understand.*, Dec. 2013, pp. 216–221.

[53] T. Y. Chai, S. S. Woo, M. Rizon, and C. S. Tan, "Classification of human emotions from EEG signals using statistical features and neural network," *International*, vol. 1, no. 3, pp. 1–6, 2010.

[54] K. Takahashi, "Remarks on emotion recognition from multi-modal bio-potential signals," in *Proc. IEEE Int. Conf. Ind. Technol. (ICIT)*, vol. 3, Dec. 2004, pp. 1138–1143.

[55] K. Ansari-Asl, G. Chanel, and T. Pun, "A channel selection method for EEG classification in emotion assessment based on synchronization likelihood," in *Proc. 15th Eur. Signal Process. Conf.*, Sep. 2007, pp. 1241–1245.

[56] R. Horlings, D. Datcu, and L. J. M. Rothkrantz, "Emotion recognition using brain activity," in *Proc. 9th Int. Conf. Comput. Syst. Technol. Workshop PhD Students Comput.*, 2008, Art. no. 6.

[57] S. J. Redmond and C. Heneghan, "Cardiorespiratory-based sleep staging in subjects with obstructive sleep apnea," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 3, pp. 485–496, Mar. 2006.

[58] G. Rodríguez-Bermúdez, P. J. García-Laencina, and J. Roca-Dorda, "Efficient automatic selection and combination of eeg features in least squares classifiers for motor imagery brain–computer interfaces," *Int. J. Neural Syst.*, vol. 23, no. 4, p. 1350015, 2013.

[59] S.-H. Oh, Y.-R. Lee, and H.-N. Kim, "A novel EEG feature extraction method using Hjorth parameter," *Int. J. Electron. Elect. Eng.*, vol. 2, no. 2, pp. 106–110, 2014.

[60] M. Khan, S. I. Ahamed, M. Rahman, and R. O. Smith, "A feature extraction method for realtime human activity recognition on cell phones," in *Proc. 3rd Int. Symp. Qual. Life Technol. (isQoLT)*, Toronto, ON, Canada, 2011, pp. 1–6.

[61] L. I. Aftanas, N. V. Lotova, V. I. Koshkarov, and S. A. Popov, "Non-linear dynamical coupling between different brain areas during evoked emotions: An EEG investigation," *Biol. Psychol.*, vol. 48, no. 2, pp. 121–138, 1998.
[62] Q. Wang, O. Sourina, and M. K. Nguyen, "EEG-based 'serious' games design for medical applications," in *Proc. Int. Conf. Cyberworlds (CW)*, Oct. 2010, pp. 270–276.
[63] E. Kroupi, A. Yazdani, and T. Ebrahimi, "EEG correlates of different emotional states elicited during watchingmusic videos," in *Proc. Int. Conf. Affect. Comput. Intell. Interact.*, Memphis, TN, USA: Springer, 2011, pp. 457–466.
[64] R. J. Davidson, "Affective neuroscience and psychophysiology: Toward a synthesis," *Psychophysiology*, vol. 40, no. 5, pp. 655–665, 2003.
[65] R. J. Barry, A. R. Clarke, S. J. Johnstone, C. A. Magee, and J. A. Rushby, "EEG differences between eyes-closed and eyes-open resting conditions," *Clin. Neurophysiol.*, vol. 118, no. 12, pp. 2765–2773, 2007.
[66] J. A. Onton and S. Makeig, "High-frequency broadband modulations of electroencephalographic spectra," *Frontiers Hum. Neurosci.*, vol. 3, p. 61, Dec. 2009.
[67] S. Akselrod, D. Gordon, F. A. Ubel, D. C. Shannon, A. C. Berger, and R. J. Cohen, "Power spectrum analysis of heart rate fluctuation: A quantitative probe of beat-to-beat cardiovascular control," *Science*, vol. 213, pp. 220–222, Jul. 1981.
[68] R. Storn, "On the usage of differential evolution for function optimization," in *Proc. North Amer. Fuzzy Inf. Process.*, Jun. 1996, pp. 519–523.
[69] K. Price, R. M. Storn, and J. A. Lampinen, *Differential Evolution: A Practical Approach to Global Optimization*. Berlin, Germany: Springer, 2006.
[70] A. K. Qin, V. L. Huang, and P. N. Suganthan, "Differential evolution algorithm with strategy adaptation for global numerical optimization," *IEEE Trans. Evol. Comput.*, vol. 13, no. 2, pp. 398–417, Apr. 2009.
[71] B. Nakisa and M. N. Rastgoo, "A survey: Particle swarm optimization based algorithms to solve premature convergence problem," *J. Comput. Sci.*, vol. 10, no. 9, pp. 1758–1765, 2014.
[72] J. Vesterstrom and R. Thomsen, "A comparative study of differential evolution, particle swarm optimization, and evolutionary algorithms on numerical benchmark problems," in *Proc. Congr. Evol. Comput. (CEC)*, vol. 2, 2004, pp. 1980–1987.
[73] R. Rakshit, V. R. Reddy, and P. Deshpande, "Emotion detection and recognition using HRV features derived from photoplethysmogram signals," in *Proc. 2nd workshop Emotion Represent. Modelling Companion Syst.*, 2016, Art. no. 2.
[74] G. Udovičić, J. ÃŘerek, M. Russo, and M. Sikora, "Wearable emotion recognition system based on GSR and PPG signals," in *Proc. 2nd Int. Workshop Multimedia Pers. Health Health Care*, 2017, pp. 53–59.
[75] Y. Xu, I. Hübener, A.-K. Seipp, S. Ohly, and K. David, "From the lab to the real-world: An investigation on the influence of human movement on emotion recognition using physiological signals," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops (PerCom Workshops)*, Mar. 2017, pp. 345–350.
[76] M. Ragot, N. Martin, S. Em, N. Pallamin, and J.-M. Diverrez, "Emotion recognition using physiological signals: Laboratory vs. wearable sensors," in *Proc. Int. Conf. Appl. Hum. Factors Ergonom.*, 2017, pp. 15–22.
[77] P. Ackermann, C. Kohlschein, J. A. Bitsch, K. Wehrle, and S. Jeschke, "EEG-based automatic emotion recognition: Feature extraction, selection and classification methods," in *Proc. IEEE 18th Int. Conf. e-Health Netw., Appl. Services (Healthcom)*, Sep. 2016, pp. 1–6.
[78] F. Feradov and T. Ganchev, "Detection of negative emotional states from electroencephalographic (EEG) signals," *Annu. J. Electron.*, vol. 8, pp. 66–69, 2014.

BAHAREH NAKISA received the B.Sc. degree in software engineering from the Azad University of Najafabad, Iran, in 2008, and the M.Sc. degree in computer science from the National University of Malaysia, Malaysia, in 2014. She is currently pursuing the Ph.D. degree with the School of Electrical Engineering and Computer Science, Queensland University of Technology. Her research interests include emotion recognition, electroencephalogram signal processing, machine learning, and computer vision.

MOHAMMAD NAIM RASTGOO received the B.Sc. degree in software engineering from Payame Noor University, Iran, in 2009, and the M.Sc. degree in artificial intelligence from the National University of Malaysia in 2014. He is currently pursuing the Ph.D. degree with the School of Electrical Engineering and Computer Science, Queensland University of Technology, Australia. His current research interests include signal processing, intelligent transportation system, machine learning, and deep learning for intelligent transportation system.

ANDRY RAKOTONIRAINY is currently the Deputy Director of the CARRS-Q and Intelligent Transport System Human Factor Research Program with CARRS-Q. He has 20 years of research and management experience in computer science and brings advanced expertise in road safety and ITS. He has established the CARRS-Q Advanced Driving Simulator Laboratory, which was the most advanced driving simulator facility in Australia. He has been proactive in investigating the use of existing and emerging ITS from multiple disciplines. It incorporates disciplines such as computer science, mathematics, human factors, engineering, psychology, and sociology. His research has made extensive use of driving simulators, traffic simulators, and instrumented vehicles for developing system prototypes, assessing cost benefits, understanding human errors, and evaluating system deployment. His research on ITS has received numerous competitive grants and generated extensive interest from road safety stakeholders.

FREDERIC MAIRE received the M.Sc. degree in pure mathematics and computer science engineering in 1989 and the Ph.D. degree in discrete mathematics from the Universite Pierre et Marie Curie, Paris 6, France, in 1993.

He is currently a Senior Lecturer with the School of Electrical Engineering and Computer Science, Queensland University of Technology, Brisbane, Australia.

His research interests include computer vision and robotics.

VINOD CHANDRAN (M'90–SM'01) received the B.S. degree in electrical engineering from IIT Madras, Chennai, the M.S. degree in electrical engineering from Texas Tech University, and the M.S. degree in computer science in 1985 and the Ph.D degree in electrical and computer engineering from Washington State University in 1985 and 1990, respectively. He is currently an Adjunct Professor with the Queensland University of Technology, Australia. He has supervised 14 Ph.D. students as the Principal Supervisor to completion and has authored or co-authored over 170 journal and conference papers. His research contributions span signal processing, image processing, and pattern recognition with applications to biometrics and biomedical systems.