This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3560362

IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2024.Doi Number

A novel partitioned random forest method based Facial Emotion Recognition

Hanif Heidari[1], M Murugappan[2,3*], Javeed Shaikh-Mohammed[4], and Muhammad E. H. Chowdhury[5*]

[1]Department of Applied Mathematics, Damghan University, Damghan, Iran.
[2]Intelligent Signal Processing (ISP) Research Lab, Department of Electronics and Communication Engineering, Kuwait College of Science and Technology, Block 4, Doha, Kuwait, 13133.
[3]Department of Electronics and Communication Engineering, Faculty of Engineering, Vels Institute of Sciences, Technology, and Advanced Studies, Chennai, Tamilnadu, India
[4]Department of Biomedical Technology, College of Applied Medical Sciences, Prince Sattam bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia
[5]Department of Electrical Engineering, Qatar University, Doha 2713, Qatar.

Corresponding author: M Murugappan (m.murugappan@kcst.edu.kw), Muhammad E.H.Chodhury (mchowdhury@qu.edu.qa)

The open access publication of this work is supported by Qatar National Library (QNL).

ABSTRACT Facial emotion recognition (FER) has been applied to various sectors, including e-learning, marketing, humanoid robot design, HMI/HCI applications, and medicine. The rapid development of intelligent technologies has led researchers to strive to improve facial emotion recognition techniques. A range of machine learning (ML) methods can be used to recognize facial expressions based on data from small to large datasets. Random Forest (RF) is simpler and more efficient than other ML algorithms. Some modified versions of RF have been developed to improve classification accuracy in the literature. Most improved RF versions modify attribute selection processes or combine them with other machine learning algorithms, increasing their complexity. Identifying an appropriate training dataset and determining its size remain open questions. The partitioned random forests (PRFs) approach is proposed as a modified strategy for improving FER. The proposed method divides multiple regions (different data lengths) into the feature space, allowing the algorithm to learn more complex decision boundaries. Using three statistical measures Lyapunov exponents (LE), Correlation Dimension (CD), and approximate entropy (AE), we evaluated the performance of machine learning algorithms over different data lengths. A crucial role for classification accuracy is played by the Lyapunov exponent or LE and the size of the dataset. A PRF is more effective on smaller datasets and with higher LE values. The proposed method for partitioning the datasets has been successfully tested on the FER dataset to classify six basic emotions (sadness, anger, fear, surprise, disgust, and happiness). Based on our numerical results, PRF performed better than traditional RF and other ML methods for FER, providing 98.37% mean absolute accuracy. Thus, a robust and useful method for improving classification rates is proposed for both small and large datasets.

INDEX TERMS Partitioned Random Forest, Random Forest, Machine Learning, Facial Emotion Recognition, Classification, Emotions.

I. INTRODUCTION

Automation and interconnectedness among humans and machines constitute the fourth industrial revolution (Industry 4.0). Due to Industry 4.0's rapid development and the impact of digitalization on our lives, large datasets are generated continuously, and their analysis has become an important topic. Human communication and decision-making are both significantly influenced by emotion. Affective computing aims to create computational systems that can recognize and react to human emotions in light of the current growth of human-computer interaction (HCI). Emotion detection is among the hottest topics in Industry 4.0, attracting decision-makers' attention. Emotion detection has been widely used in marketing, e-learning, surveillance, security, health analytics, etc.

Emotion recognition has been accomplished using a variety of inputs such as speech, body gestures, and facial expressions. For example, Hassouneh et al. [1] used electroencephalograms (EEG) and facial landmarks (virtual markers) to detect emotions in a real-time environment using machine learning (ML) methods and deep neural networks

(DNN). Batbaatar et al. [2] investigated emotion recognition from text using bidirectional Long-Short term memory (BiLSTM) and convolution neural network (CNN). Schuller et al. [3] used the hidden Markov model (HMM) for emotion recognition from speech. Similarly, Khalil et al. [4] examined the accuracy of deep learning (DL) methods in speech emotion recognition. Furthermore, Koduru et al. [5] demonstrated feature extraction to improve speech emotion recognition accuracy using Mel frequency cepstral coefficients (MFCC), discrete wavelet transform (DWT), pitch, energy, and zero crossing rate (ZCR) algorithms. Zepf et al. [6] investigated driver emotion recognition methods based on a variety of inputs for intelligent vehicle design. Recently, Maithri et al [7] examined the efficacy of artificial intelligence (AI) in automating the process of emotion recognition. It was noted that AI methods perform well under controlled conditions, while their performance gets reduced in a real-time environment.

It's noteworthy that verbal communication constitutes only one-third of human interaction, while the remaining two-thirds is non-verbal [8]. The recognition of facial expressions (FE) can be considered as one of the most essential and standard methods of non-verbal communication. The detection of facial emotions can be applied to many applications. For example, it can enhance communication between humans and robots, improve marketing science by detecting satisfying emotions, and improve distant learning methods. Also, for human-computer interactions such as clinical treatment and behavioral description, facial emotion recognition (FER) is important. Readers are referred to a latest review [9] for insights regarding emotion models employed, devices used, and classification techniques for automated visual emotion recognition. Facial emotion's effectiveness and wide applications, coupled with the emergence of Industry 4.0, make it a promising topic from a theoretical and practical perspective.

Facial emotion recognition is a subfield of social signal processing and is applied in a wide variety of areas, specifically for human and computer interaction. Various FER datasets are commonly used to evaluate facial expression recognition methods, including AffectNet, CK+, FER2013, and JAFFE. AffectNet and FER2013 are large-scale, "in-the-wild" datasets collected under uncontrolled conditions, whereas CK+ and JAFFE are smaller, controlled datasets with more standardized settings [10], [11]. Nathani [10] investigated the applicability of CNN transfer learning between different facial emotion datasets. He analyzed the FER 2013 and AffectNet datasets, designating them as the source and target datasets, respectively. His findings indicated a reduction in accuracy when using the AffectNet dataset, attributing this decline to its greater diversity compared to FER 2013. Zaman et al. [11] considered e JAFFE, CK+, FER-2013, AffectNet, and custom-developed datasets used for driver emotion recognition. Their results show that the customized dataset outperforms the considered dataset which emphasizes the effect of considering suitable datasets in efficiency of machine learning algorithms. Fard et al. [12] introduced the concept of soft-labeling which considers the occurrence probability of each emotion as its new label. They applied it on AffectNet and the dataset with the new labels was called AffectNet+.

Naga et al. [13] reviewed different methods and datasets available for facial emotion recognition (FER). Argaud et al. [14] studied methods for detecting facial expressions in Parkinson's disease (PD) patients. Slimani et al. [15] applied local binary-based methods for recognizing emotion in facial expressions. An approach to learning deep features using dense convolutional networks for FER was introduced by Sang et al. [16]. The FER system developed by Akhand et al. [17] used transfer learning and CNN. Rathour et al. [18] applied DL to recognize facial emotions from medical devices on the Internet of Things (IoT) to improve healthcare users. In other examples, Alreshidi and Ullah [19] proposed a method of extracting neighbourhood differences from features for FER. Canal et al. [20] analyzed classical and neural network models for FER. Haghpanah et al. [21] used facial landmarks and neural networks for real-time emotion recognition. A study by Lakshmi and Ponnusamy [22] used the histogram of oriented gradients (HoG) and local binary pattern (LBP) methods on detected faces to recognize emotion. In another study, Graumann et al. [23] investigated the effect of stress on patients with borderline disorders in relation to FER. Jain et al. [24] introduced a random walk (RW) and active shape model (ASM) for FER. Recently, Murugappan and Mutawa [25] introduced several geometrical features for FER. They compared the FER accuracies of support vector machines (SVM), decision trees (DT), K-nearest neighbour (KNN), extreme learning machine (ELM), and random forest (RF) methods in classifying six basic emotions, including happiness, sadness, surprise, fear, disgust, and anger. The authors found that the RF classifier outperformed other classification methods. A recent study used the VGGNet architecture with no additional training data [26] and obtained the FER2013 dataset's highest single-network classification accuracy of 73.28%. Dirik et al. considered 19 facial landmarks and a Type-2 fuzzy interface system for detecting facial emotions [27]. They achieved an accuracy of 86.17% by applying their proposed method. Recently, Dirik introduced a hybrid ANFIS-PSO method for FER and obtained 99.6% accuracy on the MUG dataset [28].

Facial emotion recognition has been an active area of research for several decades. Early approaches relied on hand-crafted features and traditional machine learning algorithms such as support vector machines and neural networks. More recently, deep learning methods have shown promising results in this field. Different types of deep learning architectures have been used for FER [29]. However, these methods often require large amounts of training data and can be computationally expensive. Zhang et al. proposed a Random Quad-Tree based ensemble algorithm (R-QT) to address the small sample size problem. R-QT enlarged the training data to obtain more diverse base classifiers [30]. A double-channel occlusion perceptron neural network model was proposed to address the issue of low face detection accuracy under complex occlusion conditions [31]. To form an occlusion perceptron neural network, the area occlusion judgment unit was designed and integrated into the VGG16 network. The perceptual neural network then extracted the features of unoccluded and less occluded regions in facial images. To reduce overfitting caused by insufficient training data samples, the transfer learning algorithm was used to pretrain parameters of the convolution layer. After optimizing the residual network, the face features of the occluding perceptron neural network and the residual network were weighted and fused.

Despite numerous studies exploring various classification techniques, the effects of dataset size on classification accuracy and optimal dataset size determination remain open issues [32]. The size of the training dataset is not a critical factor in classification problems with homogeneous classes. On the other hand, large training datasets are necessary for datasets with high variability. For example, Oyedare and Park [33] examined the effect of training dataset size on transmitter classification problems; they found that when the size of a dataset increases, classification accuracy follows a monotone increase and convergence sequence. Chu et al. [34] studied feature selection in Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET) to detect early Alzheimer's disease. Their results indicated that larger datasets lead to more accurate results, although some minor fluctuations still exist. The effects of dataset size on logistic regression (LR) and neural network (NN) models were studied by Baillyet al. [35]. They showed that dataset size does not affect the accuracy of LR and deep NN methods. Althnian et al. [32] examined the effect of medical dataset size on the accuracy of NN, nave Bayes (NB), SVM, DT, and RF classifiers. Their results showed that SVM, ANN, and RF are appropriate classification methods for small-size datasets. Therefore, RF is primarily used for classification tasks with small data sets. In this study, we investigate RF classifiers for facial emotion recognition. Also, we investigate the effect of dataset size on the RF classifier.

In the previous work [25], the RF classifier achieved a maximum mean emotion recognition rate of 98.17% (number of trees N=300) based on the Inner Circle Area of the Triangle (ICAT). Other triangular features, including the Area of the Triangle (AoT) and the Inner Circle Circumference of the Triangle (ICCT), also yielded maximum mean emotion recognition rates of 97.97% (N=300) and 97.47% (N=55). In contrast, when all the above three features were combined and compared to other classifiers such as SVM, PNN, ELM, DT, and KNN, the RF classification could achieve a maximum mean classification rate of 98.05% (N=620). The number of trees (N) in the RF classifier was heuristically varied between 50 and 3000, with increments of 10. Even though the RF classifier showed a higher mean emotion recognition rate than the SVM, DT, KNN, PNN, and ELM classifiers, it required the highest computation time. The maximum computation times for AoT, ICAT, and ICCT features were 31573 s, 34471 s, and 35263 s, respectively, on a Windows 10 machine equipped with an Intel i7 processor at 2.43GHz and 32 GB of memory.

A large sample size is one of the main reasons for requiring more computation time. Literature indicates that a system's computational power is directly related to the input size in ML problems. Therefore, the current work aims to develop an intelligent algorithm for the RF classifier to achieve a higher emotion recognition rate with a lesser computation time. We propose a novel partitioning method algorithm to reduce the sample size without losing any information for classifying emotions. To the best of our knowledge, almost all existing research discusses the need for a larger dataset to produce better results. Despite previous studies, this study found that reducing the dataset improved accuracy when using an RF method. Additionally, this work provides a confidence value for each classification result, which is very meaningful from a theoretical and practical perspective. This work uses 190967 samples and 25 features for FER. Thus, the dataset size is large enough to eliminate overfitting issues [36], [37], [38], [39].

The main contribution of this manuscript is threefold.

- First, this paper proposes a modified version of RF that significantly increases classification accuracy in real-time emotion recognition using virtual markers.
- Second, the effects of dataset length, Lyapunov exponents (LE), Correlation Dimension (CD), and approximate entropy (AE) on classification accuracy are investigated.
- Thirdly, we compared the performance of the proposed PRF with other ML algorithms, conventional RF classifiers and with open-source FER dataset for benchmarking.

Overall, real-time FER is an important area of research with numerous applications in fields such as psychology, marketing, online learning, virtual reality, and security [1], [18], [21], [40]. Despite significant progress in this field, accurately recognizing real-time emotions from facial expressions remains a challenging task due to the need to work with limited data and operate quickly [18], [40]. In this paper, we propose a novel partitioned random forest method for facial emotion recognition. Our approach leverages the strengths of random forest algorithms while introducing a partitioning scheme that improves the accuracy of emotion recognition.

II. RELATED WORKS

Random forest (RF) is a machine learning algorithm that uses supervised learning methods and can be applied to both

IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

classification and regression problems [41]. It operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. The algorithm has three main hyperparameters: node size, number of trees, and number of features sampled. Random forest is an ensemble learning method and is considered to have good scalability and parallelism to high-dimensional data in classification [42]. The random forest method is a powerful machine-learning algorithm that has been applied across several industries to solve complex problems and make better decisions [43], to make predictions and identify patterns in data.

Random forest algorithms have also been used for facial emotion recognition. For example, one study proposed and developed a methodology to identify facial emotions using facial landmarks and a random forest classifier [37]. Faces were first identified in each image using a histogram of oriented gradients with a linear classifier, image pyramid, and sliding window detection scheme. Then, the random forest algorithm was used as the facial expression classifier. Random Forest is a type of ensemble learning, where multiple models are combined to improve performance. It can be used with other machine learning algorithms to improve accuracy and reduce overfitting. For example, a study used a combination of random forest and convolutional neural network for facial expression recognition [42]. Based on the discriminative representation of features, a recent study proposed a DNA-RCNN (Deep Normalized Attention-based Residual Convolutional Neural Network) to extract the appropriate features [44]. The proposed M-RF (modified-random forest) with an empirical loss function was used for classification. The learning weights on the data subset reduce loss between the predicted value and the ground truth, allowing for more precise classification. According to Dapogny et al., FER is a spatiotemporal event, and a pairwise conditional random forest was introduced to address this issue [45][40]. Gharsalli et al. used RF to select features from the FER dataset in their study [46].

Random forests have been successfully applied to a wide range of practical problems, but they are not as accurate or computationally efficient when applied to large datasets [47], [48], and [49]. Shahhosseini and Hu proposed a weight optimization method for increasing RF accuracy [50]. In [51], the authors have combined the adaboost classifier with RF for facial emotion classification and achieved a maximum mean accuracy of 92.5% using the CK+ dataset. In a recent study, the combination of RF with a Deep Neural Network (Convolutional Neural Network (CNN)) was proposed by Arnaoud et.al to classify facial expressions. Their proposed CNN-RF achieved a maximum mean classification rate of 86.6% in the CK+ dataset. In their study, Yin et al. suggested reducing computational time by using Spark and parallel RF [49]. A disjoint subset of attributes and datasets was taken into account by Kulkarni and Sinha to improve RF accuracy [52]. The dataset is divided into a number of considered trees based on their approach. The elements of each partition that contain subsets of features are used to train its related tree. In their methods, the elements of each partition are randomly selected, and each partition is of the same size. Finding the optimal dataset size remains the most challenging step to improving RF accuracy.

To our knowledge, finding the ideal dataset size for improving RF accuracy remains an open issue. A partitioned random forest (PRF) method is proposed in this paper as a way to improve random forest algorithms' accuracy. In these methods, the feature space is partitioned into multiple regions so the algorithm can learn more complex boundaries for decision-making.

In this paper, partitioned random forest (PRF) methods have been proposed to improve the accuracy of random forest algorithms. These methods introduce a partitioning scheme that divides the feature space into multiple regions, allowing the algorithm to learn more complex decision boundaries.

III. MATERIALS AND METHODS

A. DATA ACQUISITION

This work used the FER database developed by Murugappan et al. [25] to recognize facial emotions. This study involved 85 participants mean aged 24.5 years (21 – 32 years) who participated voluntarily, with a sex ratio of 55: 30 (male: female). All the participants are university undergraduate students. The investigation was conducted in a controlled environment (room temperature at 25ºC) with different backgrounds (solid black and advertisements) to acquire facial expression data for six emotions (happiness, sadness, anger, fear, disgust, and surprise) with a camera on an Apple MacBook with a resolution of 2560 × 1600 at 30 frames per second (FPS).

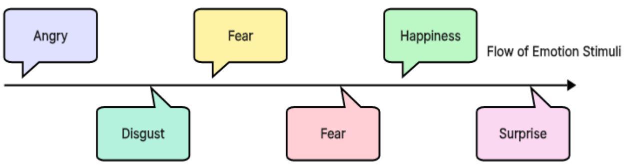


FIGURE 1. Emotion elicitation protocol used for facial expression recognition.

The participants were presented with an automated PowerPoint slide presentation to elicit their emotional responses and asked to express each emotion ten times. An overview of the emotion elicitation protocol used here to acquire six samples of facial expressions is shown in Figure 1. The figure also shows the order and label of emotions used in this study. A 10s break was given to the participants to avoid any feedback from previous emotion elicitation. At the end of each trial, the participants will be asked to report the emotion they felt during the expression. Eight virtual markers were

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3560362

placed on defined locations (as shown in Figure 2(a)) based on a mathematical model of the participant, and the marker coordinates were used to formulate seven triangles (as indicated in Figure 2(b) and Table 1).

A more detailed description about the data acquisition environment, data acquisition device specification, marker placement, formation of seven triangles based on eight markers, and triangular features extraction is given in [22]. A feature vector of size 190967 × 25 was derived and the features were normalized using the average mean reference method. Finally, the normalized features were used to analyze the performance of the proposed partitioned random forest (PRF) method to classify facial emotional expressions.

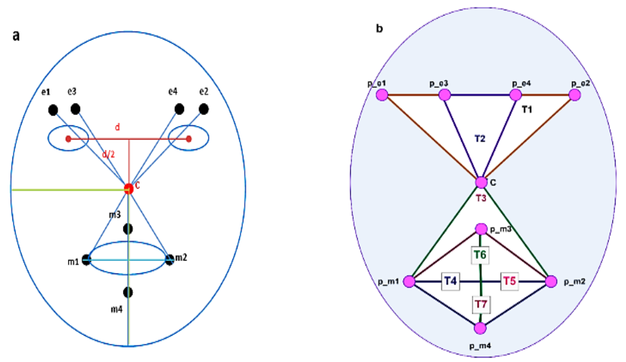


FIGURE 2. (a) Placement of virtual markers, (b) triangles used for feature extraction.

TABLE 1. Markers used to formulate the seven triangles used for feature extraction.

Triangles	Markers
T1	p_e1, p_e2, and C
T2	p_e3, p_e4, and C
T3	p_m1, p_m2, and C
T4	p_m1, p_m3, and p_m4
T5	p_m2, p_m3, and p_m4
T6	p_m1, p_m2, and p_m3
T7	p_m1, p_m2, and p_m4

B. RANDOM FOREST METHOD

The random forest method is an ensemble data mining technique that can be used for both classification and regression problems. It is an improved form of the decision tree method with desirable features, including the ability to handle high-dimensional problems, parallel processing, and reduced computational time. Such features have popularized RF in the data mining field. For example, Murugappan and Mutawa [25] used RF to detect emotions in facial images; Liu and Ge [53] used weighted RF to detect faults in industrial processes. A study by Shah et al. [54] showed that RF outperformed logistic regression and KNN for text classification.

Consider a p-dimensional dataset with N elements in dataset Z and r classes. Figure 3 illustrates the method of RF classification. The RF method splits the p features into k disjoint subsets randomly and applies the decision trees method to each subset. Each tree generates a classification result for each input. The final classification result is the class that has the highest score.

C. PARTITIONED RANDOM FOREST METHOD

Despite the great features of RF, it is inefficient for solving some practical problems due to its higher computation time and memory requirement. Improved versions of RF have been developed to overcome these difficulties. For example, Paul et al. [41] divided the features set into subsets of important and unimportant features, resulting in improved classification accuracy. It was also demonstrated that defining the upper limit of the total number of trees in the RF classifier could improve the classifier's accuracy. As the traditional RF classifier suffers from a long run time resulting in slower fault identification, Han et al. eliminated trees with low accuracy and applied a sub-forest optimization approach to improve the RF classifier prediction accuracy in small datasets by modifying the hyperparameters [55].

Algorithm 1: Random Forest method

Input: dataset Z, number of ensemble c, number of selected features *Ni*
for *i*= 1 to *c* **do**
Randomly sample the training data *D* with replacement to produce *Di*
Create a root node, *Ni* containing *Di*
Call BuildTree(*Ni*)
end for
BuildTree(N):
if *N* contains instances of only one class **then**
return
else
Randomly select x% of the possible splitting features in *N*
Select the feature *F* with the highest information gain to split on
Create f child nodes of *N*, *N*1,..., *Nf*, where *F* has *f* possible values (*F*1, …, *Ff*)
for *i*= 1 to *f* **do**
Set the contents of
*Ni*to *Di*, where *Di* is all instances in *N* that match
Fi
Call BuildTree(*Ni*)
end for
end if

FIGURE 3. The pseudocode of the random forest algorithm.

To the best of our knowledge, a handful of studies have explored the impact of dataset size on RF accuracy. In one study, Rodriguez-Galiano et al. [56] evaluated the effect of training dataset size on the classification accuracy for RF methods. They found that the RF classification error increased by shrinking the training dataset. Racz et al. [57] examined the effects of dataset size and train/test split ratio on the classification accuracy of Xtreme Gradient Boosting (XGBoost), NB, SVM, feedforward neural networks (FNN), and PNN. Their results indicated that XGBoost was relatively unaffected by the dataset size, whereas all the other methods considered were more efficient when the dataset size increased. Catal and Diri [58] used hypotheses tests in

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3560362

examining the effect of dataset size on the accuracy of software fault prediction. They found that both RF and NB classifiers were appropriate for large and small datasets.

In random forest-based machine learning training, partitioned random forests address an important and complex challenge by optimizing the dataset size [59]. Small training datasets may result in underfitting, while large training datasets may result in overfitting. An ECG time series classification was performed by Gupta et al. [60] using 20 samples per class. Shahinfar et al. investigated the impact of training dataset size on classification accuracy in project-specific camera trap models. For each class, 150-500 images are sufficient to train deep-learning models [61]. By Goudjil et al. [62], the SVM was applied to text classification with 20 samples for each class. Based on our assessment, the results obtained in the above references are only valid for the datasets considered in those references and cannot be generalized. In the ImageNet-Sketch dataset [63], there are 1000 classes, and 500 samples in each class are not sufficient to train deep learning algorithms, while 100 samples in each class may cause overfitting. Therefore, the number of samples per class should depend on the number of classes and should not be fixed. The number of training samples for each class is averaging between nn^2 and $6{*}nn^2$ where nn is the number of classes. Binary classification typically requires four to 24 samples per class. It may be necessary to use 1000000 to 6000000 samples for a dataset with 1000 classes depending on the correlation between the data in each class. A small number of samples is required if there is a high correlation between the elements of a class, whereas large samples are required if the correlation is low. Across different training dataset sizes, the best accuracy is determined by the size of the training dataset used as the optimal training dataset. Testing must be conducted on a large enough set of data to ensure the results are robust. Several famous datasets, including MNIST-10 [64], which contains 10000 test elements, satisfy the condition. For small datasets or datasets in which the test and train sets are not distinguished, we suggest considering the $(1+nn)^2$ to $6{*}(1+nn)^2$ elements with a 20-80 or 30-70 test/train ratio split. A model should be run at least three times independently to ensure stability where the train and test sets are independent without any intersection.

Our investigations indicate that one can select an optimal dataset size for each classification problem to maximize the RF classification accuracy. For a large training dataset, it is important to consider RF with a deep structure to ensure acceptable results. In addition, RF may fall into overfitting for large training datasets. For a small training dataset, underfitting may occur, leading to inaccurate results. In this paper, we aim to show that there exists an optimal training dataset that maximizes RF accuracy. In addition, we will provide an algorithm for finding the optimal training dataset. To design a proper algorithm, it is not appropriate to ignore some data as it could potentially have vital information that is left out, impacting the accuracy of results outside of the sample. Therefore, this work proposes a partitioning method to decompose the whole dataset into parts with appropriate lengths. The RF method can be implemented on each subset in parallel, preventing the need for additional run-time costs. Normalized data is used to construct partitioned subsets, and a norm-1 partitioning method is used for each element. The partitioned random forest (PRF) algorithm is illustrated in Figure 4. In Step 1 to Step 3, based on the newly proposed subject independent feature (f_{26}), the partitioned random forest method tries to split the dataset into some other smaller datasets. As it mentioned earlier, the number of elements in each partition should be between nn^2 to $6{*}nn^2$ if dataset is large enough or the train and test set are separated already. Otherwise, the dataset set should be partitioned into subsets with the number of elements between $(1+nn)^2$ and $6{*}(1+nn)^2$. In this case, 70-30 train/test split ratio is used for facial emotion classification. Here, we used subject-independent methods to split the training and testing set and ensured that the data in both sets does not belong to the same subject.

To partition the dataset, PRF generates a new feature, which is the sum of previous features (Step 2 in Algorithm 2). The value of the new feature is used to detect the element of each partition. An element belongs to i^{th} partition if the value of the new feature lies between *i-1* and *i* (Step 3 in Algorithm 2). If the classification accuracy of a partition is not good, we divide the partition into some other sub-partitions to improve the accuracy. Step 5 in Algorithm 2 describes the relationship between the number of elements in sub-partitions and the number of features.

Algorithm 2: Partitioned random forest method
Inputs: Dataset Z, desired accuracy ACC, Confident level (95%)
Step 1: Normalize Z (Described in the next section)
Step 2: For each n∈Z, add the feature $f_{26}(n) = \sum_{k=1}^{25} f_k(n)$
Step 3: Partition Z to Zi, i=1,…, M subsets such that Min$f_{26}(Zi)$> i-1 and Max$f_{26}(Zi)$< i.
Step 4: Apply RF on each Zi and evaluate the accuracy, Ai
Step 5: **for i=1 to M**
If accuracy Ai<ACC and (the number of elements) >*nn* then **S=0, i=1** **while S<ACC and i<=6**
(a) Partition Ai in some disjoint subsets with $i{*}(1+nn)^2$ elements,
(b) evaluate the accuracy at each subset
(c) S=max(the mean of subsets classification accuracy as the classification accuracy of Zi, (Ai))
(d) Ai=S
(e) i+=1
End while
End for
Step 7: Print the average of (Ai) as the accuracy of the model

FIGURE 4. **The pseudocode of partitioned random forest (PRF) algorithm.**

D. PERFORMANCE METRICS

In this work, the performance of the classifiers is measured through accuracy, True Positive Rate (TPR), and False Positive Rate.

$$Accuracy\ (Acc) = \frac{\mathrm{TP+TN}}{\mathrm{TP+FP+FN+TN}} \tag{1}$$

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3560362

$$True\ Poisitive\ Rate\ (TPR) = \frac{TP}{TP+FN} \quad (2)$$

$$False\ Positive\ Rate\ (FPR) = \frac{FP}{FP+TN} \quad (3)$$

IV. RESULTS AND DISCUSSION

There are 190967 samples with 25 features used to classify facial images for emotion detection [25]. These features include the marker features (p_e1,..., p_e4, p_m1,…, p_m4) as well as the triangle features (triangles T1,…,T7).In Murugappan and Mutawa [25], six different types of ML algorithms (SVM, PNN, KNN, ELM, DT, and RF) were applied for facial emotion classification using three triangle features, namely Area of Triangle (AoT), Inner Circle Area of Triangle (ICAT), and Inner Circle Circumference of Triangle (ICCT). They showed that the RF method outperforms other methods with an accuracy of approximately 97% using the ICAT feature (Figure 5). This comparison should suffice to illustrate the benefits of the proposed PRF method.

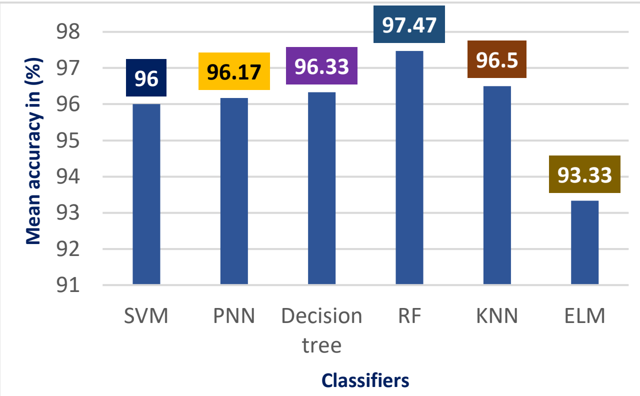


FIGURE 5. The classification accuracy of different methods in facial emotion recognition.

Our goal was to find a procedure that accurately classifies emotions with at least a 95% confidence level such that the mean accuracy is greater than the mean accuracy of the whole dataset when using RF. As a rule of thumb, we considered the train-test ratio 70-30 with 30 trees. Therefore, the train/test split ratio remains consistent between the original dataset and the partitioned datasets, ensuring that accuracy comparisons are valid. We validated our findings by applying the method three times independently on each dataset. Almost identical results were obtained, thus demonstrating the stability of the PRF method. A 96% classification accuracy was obtained for the entire dataset. Accordingly, we should choose a partition such that the mean accuracy in each partition is greater than 96%, and the accuracy of each partition must be at least 95% (confidence level specified by the user). Table 2 shows a comparison between RF and the first stage of PRF.

According to Table 2, the PRF method produces satisfactory results, except for subsets Z4, Z5, and Z6. In this context, Z refers to the whole (or complete) set of data given to the classifier, while Z1-Z9 refers to the sub-sets generated by the proposed methodology. All subsets achieve an accuracy higher than 80% and a maximum of 100% is achieved using Z1, Z8, and Z9. In addition to accuracy, the receiver-operating characteristic curves (ROC) and the area under the curve (AUC) are used to evaluate and compare the efficiency of classification algorithms [65]. ROC is the plot of sensitivity (TPR) against 100% specificity (FPR) for different threshold points. The Area Under the ROC (AUC) is a measure that determines the ability of the considered classification algorithm for distinguishing the elements between two groups. The higher AUC indicates more precise results. A classification method with AUC<0.75 is not useful, while that with AUC>0.97 is highly accurate [65]. Table 3 gives its AUC of whole data and data subsets. Here, we have obtained the same value of AUC in recognizing emotions for Z1, Z2, Z3, Z4, Z7, and Z8 and different values for complete data, Z5 and Z6. Figure 6 shows ROC for the PRF classifier of whole data, Z1 (since it's the same for Z2, Z3, Z4, Z7, and Z8), Z5, and Z6. These results confirm that PRF is highly accurate for recognizing FER, except for Z5, which has AUC<0.97 for angry and disgust emotions.

TABLE 2. The classification accuracy of the random forest method for the original dataset and the partitioned random forest subsets.

Dataset	Z	Z1	Z2	Z3	**Z4**	**Z5**	**Z6**	Z7	Z8	Z9
Number of elements	190967	5793	14817	43517	**53170**	**55357**	**12488**	4394	403	843
Accuracy %	96.35	100.00	99.98	99.96	**89.18**	**80.93**	**89.81**	99.70	100.00	100.00

TABLE 3. AUC of facial emotion recognition using PRF subsets.

	Area under curve (AUC)								
Subset/Emotion	**Complete**	**Z1**	**Z2**	**Z3**	**Z4**	**Z5**	**Z6**	**Z7**	**Z8**
Angry	0.99	1.00	1.00	1.00	1.00	0.95	0.97	1.00	1.00
Disgust	0.99	1.00	1.00	1.00	1.00	0.95	0.98	1.00	1.00
Fear	0.99	1.00	1.00	1.00	1.00	0.97	0.99	1.00	1.00
Sadness	0.99	1.00	1.00	1.00	1.00	0.98	0.99	1.00	1.00
Happiness	0.99	1.00	1.00	1.00	1.00	0.98	0.99	1.00	1.00
Surprise	0.99	1.00	1.00	1.00	1.00	0.97	0.99	1.00	1.00

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3560362

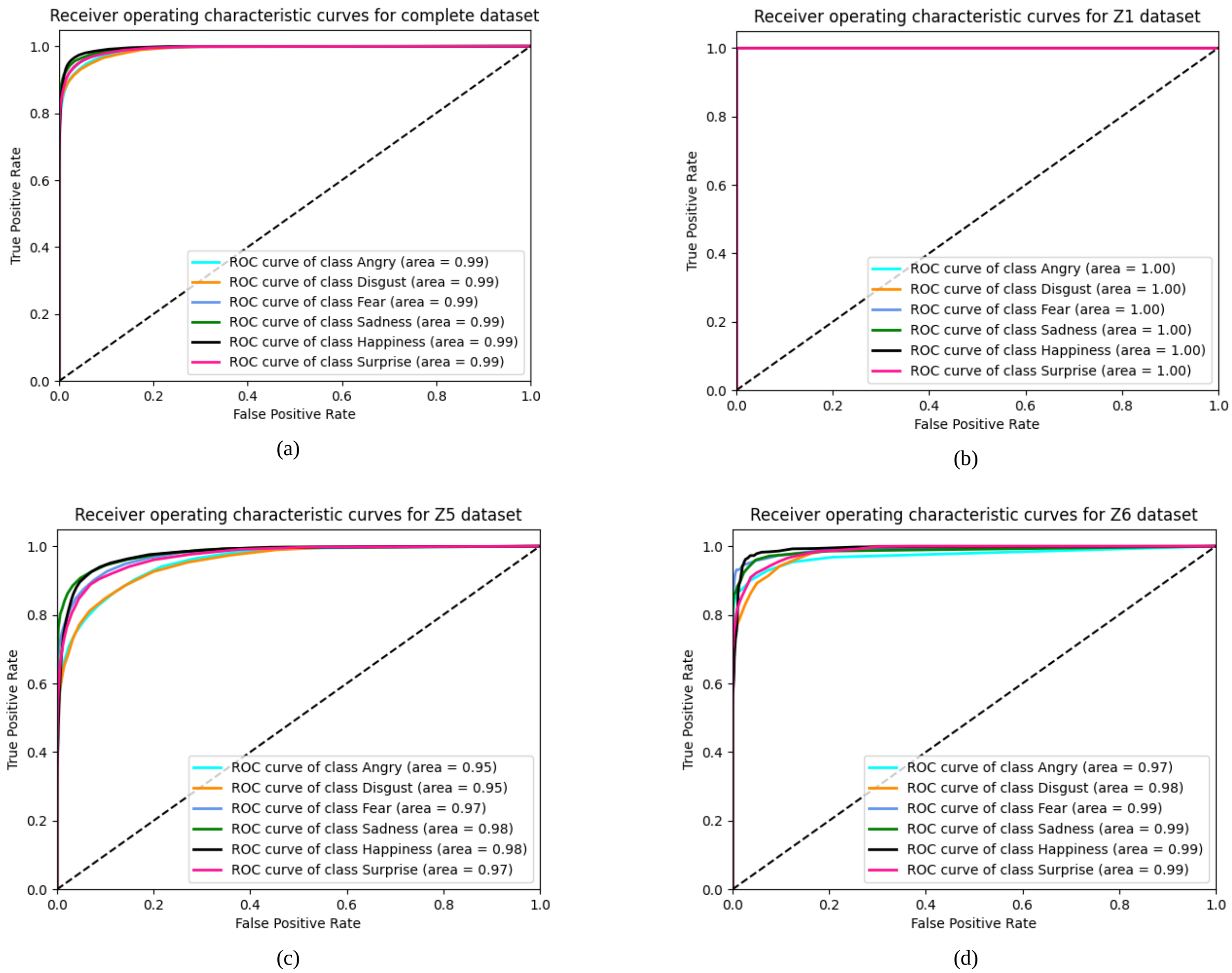


FIGURE 6. Receiver Operating Characteristics (ROC) curve of the classifiers (a) whole data (b) Z1 subset (c) Z5 subset and (d) Z6 subset.

TABLE 4. The confusion matrices for Z4, Z5, and Z6.

	Emotion	Anger	Disgust	Fear	Sadness	Happiness	Surprise
Confusion matrix for *Z4*	**Anger**	**92.20**	1.37	3.83	1.88	0.35	0.38
	Disgust	10.20	**81.90**	2.68	2.96	0.45	1.90
	Fear	6.16	0.78	**89.80**	2.90	0.07	0.26
	Sadness	4.94	0.82	3.27	**90.10**	0.12	0.76
	Happiness	4.73	1.95	1.26	1.10	**90.60**	0.32
	Surprise	4.60	1.75	1.56	1.41	0.19	**90.50**
Confusion matrix for *Z5*	**Anger**	**71.80**	5.73	6.58	4.27	7.34	4.27
	Disgust	5.78	**73.10**	3.76	2.48	7.56	7.29
	Fear	4.08	4.36	**83.30**	2.11	2.66	3.46
	Sadness	3.41	3.29	4.15	**85.00**	1.67	2.52
	Happiness	2.34	2.63	1.53	1.21	**89.70**	2.54
	Surprise	3.16	3.74	2.74	0.87	6.80	**82.70**
Confusion matrix for *Z6*	**Anger**	**79.80**	4.64	0.99	2.65	6.29	5.63
	Disgust	0.20	**83.30**	3.05	1.02	5.30	7.13
	Fear	0.17	1.88	**92.50**	0.34	0.69	4.45
	Sadness	0.00	2.18	0.94	**92.80**	0.00	4.05
	Happiness	0.00	0.65	0.00	0.11	**97.60**	1.64
	Surprise	0.18	1.50	0.80	0.35	4.24	**92.90**

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3560362

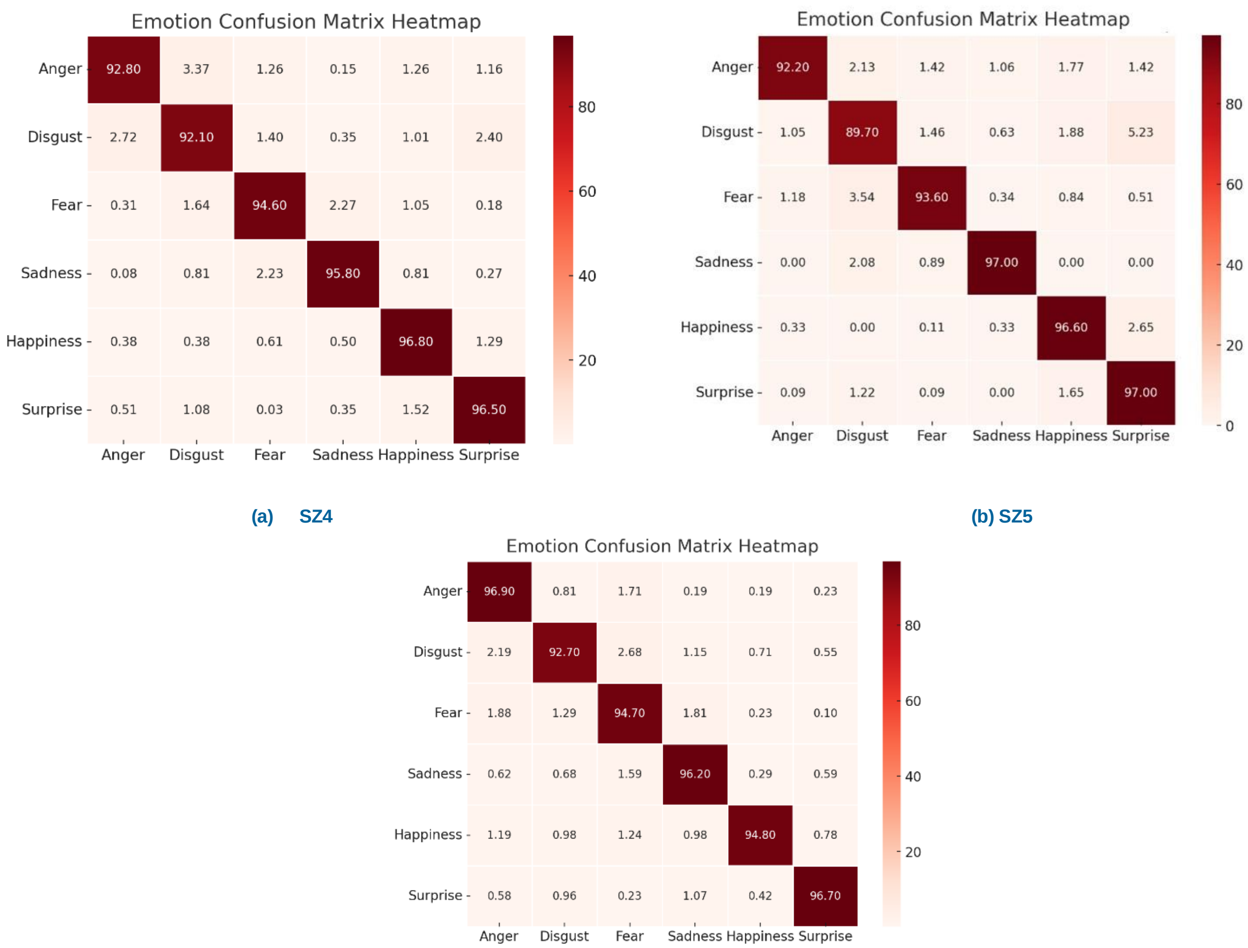
TABLE 5. The average classification accuracy for the subsets of SZ4, SZ5, and SZ6; where each subset has at most 250 elements.

Dataset	*SZ4* (*nn*=250)	*SZ5* (*nn*=200)	*SZ6* (*nn*=250)
Average ACC%	**95.60**	**95.04**	**95.13**
Min ACC %	53.33	50	64
Max ACC %	100	100	100
Median ACC%	**100**	**98.33**	**100**
Standard deviation	7.93	8.23	8.66

*ACC: Accuracy

Table 4 contains the confusion matrices for these three subsets. The confusion matrices of Z5 and Z6 indicate that anger emotion recognition is a challenging task. Although the confusion matrix of Z4 indicates maximum recognition accuracy for anger and minimum recognition accuracy for disgust, the second line shows that the low accuracy in disgust emotion recognition is caused by the algorithm's weakness in distinguishing between anger and disgust. Therefore, identifying anger emotion recognition needs further attention.

Therefore, Z4, Z5, and Z6 were divided into disjoint subsets. A set of disjoint subsets of Z4, Z5, and Z6 is denoted by SZ4, SZ5, and SZ6, respectively. Table 5 illustrates the statistical properties of the classification accuracy for each of the considered subsets. The median accuracy of SZ4, SZ5, and SZ6 was greater than 98%, suggesting that more than half of the subsets had an accuracy greater than 98%. An accuracy of less than 65% indicates that the analyzer needs to concentrate on these subsets, investigate any outlier data, or develop more samples to improve accuracy. Based on Algorithm 2, the set Z4, Z5 and Z6 should be partitioned into subsets with $i*(6+1)^2$, i=1, …, 6 elements. For simplicity without loss of generality, we consider $(6+1)^2 \cong 50$. The calculation concludes i=5, 4 and 5 for Z4, Z5 and Z6 respectively. The confusion matrices for SZ4, SZ5, and SZ6 are shown in Figure 7.



(a) SZ4

(b) SZ5

(c) SZ6

FIGURE 7. The augmented confusion matrices for (a) SZ4, (b) SZ5 and (c) SZ6.

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3560362

The results in Table 5 and Figure 7 prove that PRF outperforms existing methods for FER. The approximate entropy, correlation dimension, and Lyapunov exponent were calculated to investigate the underlying theoretical reasons for this improvement. To reduce complexity, we used MATLAB software and considered default values as initial values for these parameters' computation, i.e., in all computations, the embedding dimension and time lag were taken as 2 and 1, respectively. The expansion steps were taken in the range (1,5) for computing the Lyapunov exponent, and 10 points were considered when calculating the correlation dimension. The results for these mathematical properties are presented in Table 6.

TABLE 6. Mathematical properties of the original and partitioned datasets.

Dataset	Approximate Entropy	Correlation Dimension	Lyapunov Exponent	Number of Elements	Accuracy %
Z	1.538506	7.627926	-0.142	190962	96.35
Z1	1.150326	6.507724	0.301443	5973	100.00
Z2	1.284794	6.790684	0.336175	14817	99.98
Z3	1.466413	7.066216	0.334643	43517	99.96
Z4	1.367146	7.126334	-0.13103	53170	90.73
Z5	1.371762	7.417851	-0.12988	55357	83.57
Z6	1.101974	6.604944	0.323424	12488	96.35
Z7	1.084597	6.228043	0.324628	4394	99.70
Z8	**1.18898**	**4.471484**	**0.140159**	**403**	**100.00**
Z9	**0.540082**	**5.847646**	**0.295327**	**843**	**100.00**

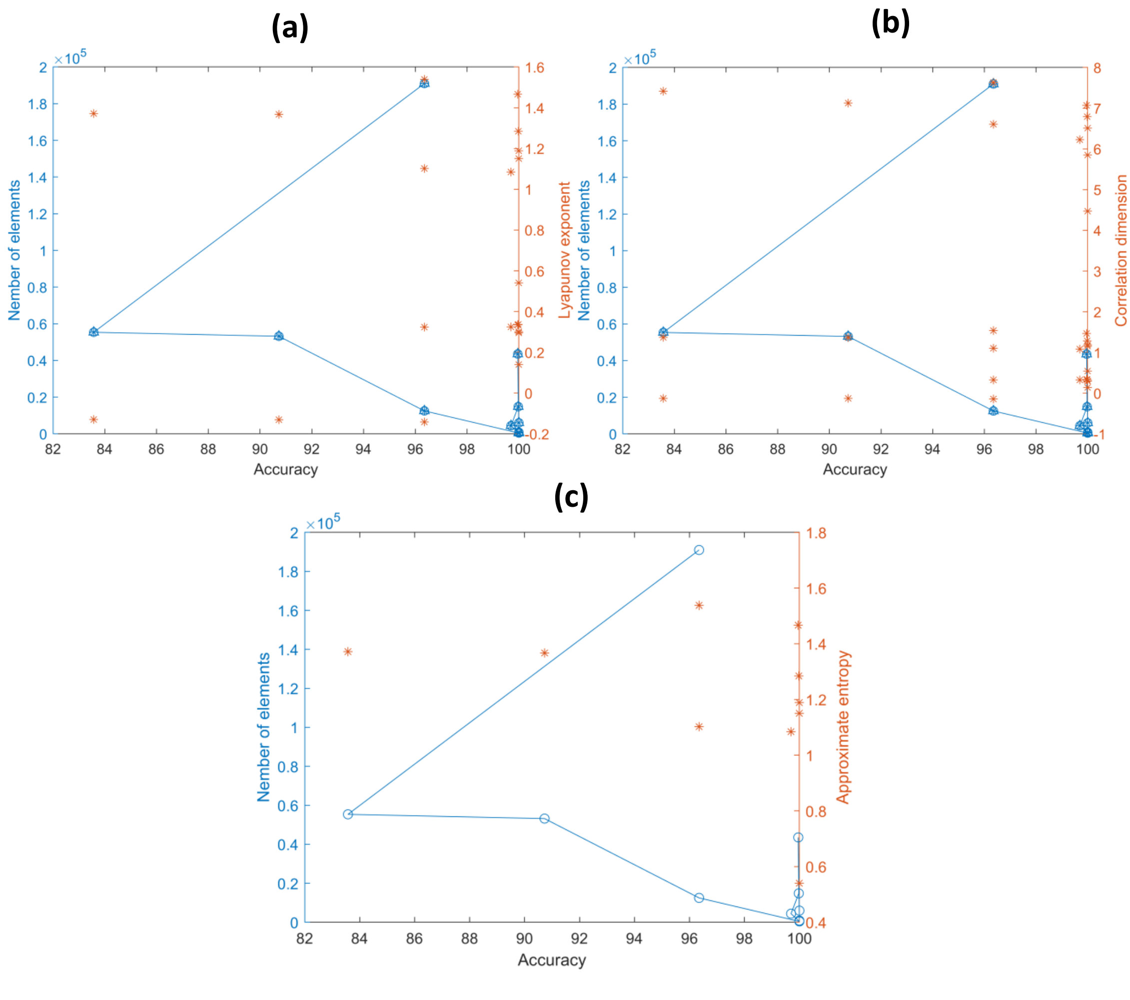


FIGURE 8. (a) Lyapunov Exponent (LE) (b) Correlation Dimension (CD) (c) Approximate Entropy (AE).

The results indicate that datasets with negative Lyapunov exponents exhibit the lowest classification accuracy. As a result, datasets with chaotic behaviors are more suitable for classification. Heidari and Velichko recently introduced a new classification method, LogNNet, by applying a chaotic matrix to input data [66]. Their results confirm the current understanding of the relationship between classification accuracy and the Lyapunov exponents. However, the Lyapunov exponent is not sufficient to determine RF classification accuracy. Consider the datasets Z6 and Z9, which have positive Lyapunov exponents. While the Lyapunov exponent Z6 is greater than the Lyapunov exponent Z9, the classification accuracy in dataset Z6 is smaller than that in dataset Z9. Similar problems are observed with datasets Z and Z5. The Lyapunov exponent of Z is smaller than that of Z5, while the classification accuracy of dataset Z5 is lower than that of Z.

According to our analysis, the number of elements is also a significant factor that affects classification accuracy. It is noteworthy that although datasets Z6 and Z7 have almost the same positive Lyapunov exponent, their classification accuracy does not match. This mismatch is due to the differences in the number of elements in the two datasets. Figure 8(a) shows a plot of accuracy concerning the number of elements and the Lyapunov exponents. Comparing classification accuracy between two distinct datasets with the same Lyapunov exponents reveals that the smaller datasets possess significantly higher classification accuracy. Similarly, comparisons were made regarding the accuracy of classification with dataset size and the properties of the dataset, including correlation dimension (Figure 8(b)) and approximate entropy (Figure 8(c)). The results suggest that a dataset with a smaller correlation dimension and approximate entropy values generates more accurate classification results in emotion recognition.

To illustrate the robustness and generality of the proposed method, we also considered the 428-landmark 3D facial images, introduced in [67]. The dataset contains 429 images for detecting happiness, surprise, sadness, and anger emotions. The feature space contains 468 three-dimensional landmarks. Therefore, the dataset contains 3*468 features, which are much larger than the number of elements in the dataset. Based on Algorithm 2, the dataset should be partitioned into subsets with the number of elements between 4^2 (= 16) and $6*4^2$ (= 64) to achieve maximum accuracy. By following Algorithm 2, data are normalized, and feature (4) is added. The data is sorted by the feature f_{1405} in ascending order.

$$f_{1405} = \sum_{i=1}^{144} f_i \tag{4}$$

Next, the sorted dataset is partitioned into subsets with 16*k elements, in which k could be 1, …, 6 to obtain maximum accuracy. We started with k=1 and increased it in steps to achieve the desired accuracy. The computation was repeated three times independently, in which 80% of the data was used for training and 20% for testing the accuracy. The mean of accuracy in different runs was considered as the accuracy of the method. The results for k=1,…, 4 are shown in Table 7, which indicates partitioning with 2*16 elements could accurately classify anger, happiness, sadness, and surprise emotions. Although the partition with 32 elements could classify the emotions exactly, we continued Algorithm 2 for investigating the trends for classification accuracy. The partitions with 32 or 48 elements could classify the emotions exactly, while the classification accuracy is decreased for *nn*=64 (see Table 7). This phenomenon confirms our claim about the existence of an optimal partition for the random forest classifier.

TABLE 7. Mean accuracies for different partition lengths and number of partitions for the 468-landmark 3D facial image dataset.

Partition length	Number of partitions	Mean accuracy (%)
427	1	97.70
16	26	97.82
32	13	100
48	8	100
64	6	98.72

The values of approximate entropy, correlation dimension, and Lyapunov exponents versus number of partitions are shown in Figures 9, 10, and 11, respectively. From Figure 9, it is seen that for *nn*=32 and *nn*=48, at which highest mean accuracy was achieved, have lower approximate entropy values than for *nn*=16 (except for the negative values for *nn*=16).

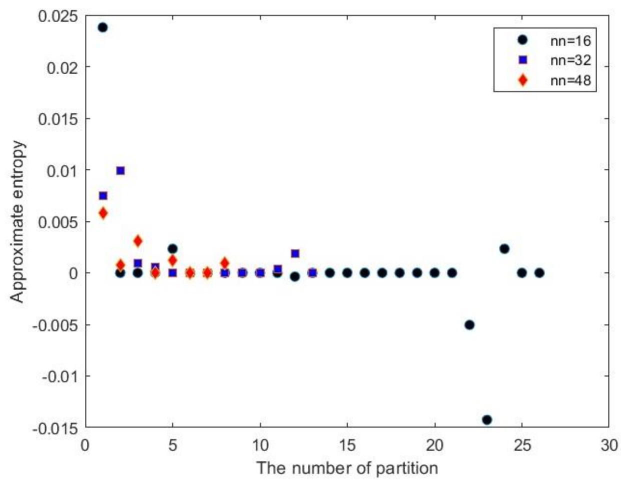


FIGURE 9. Approximate entropy versus number of partitions for the 468-landmark 3D facial image dataset.

From Figures 9 and 11, it can be observed that negative Lyapunov exponent or negative approximate entropy values cause inaccuracy in the RF method, while lower approximate entropy or higher Lyapunov exponent values lead to higher accuracy. From Figure 10, it is seen that RF generates more accurate results on datasets with higher correlation dimension values. These results are important for two reasons. First, it proves the robustness of the proposed method in small datasets, where neural networks and deep learning methods fail. Second, existing facial emotion recognition approaches have problems and give inaccuracies for distinguishing between anger and sadness or happiness and surprise. However, the proposed method could distinguish between these emotions accurately.

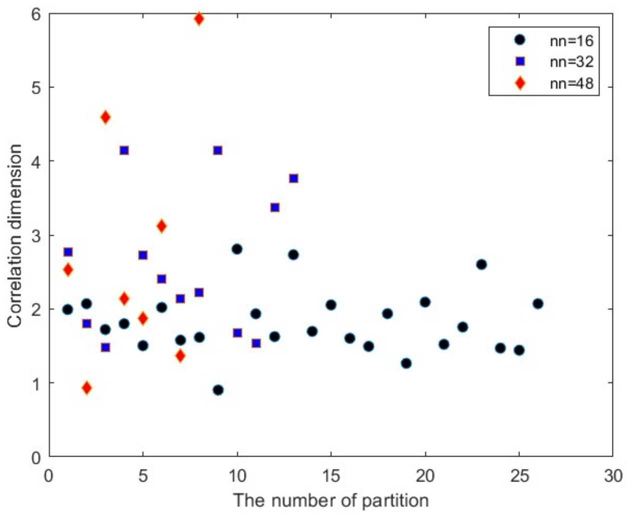


FIGURE 10. Correlational dimension versus number of partitions for the 468-landmark 3D facial image dataset.

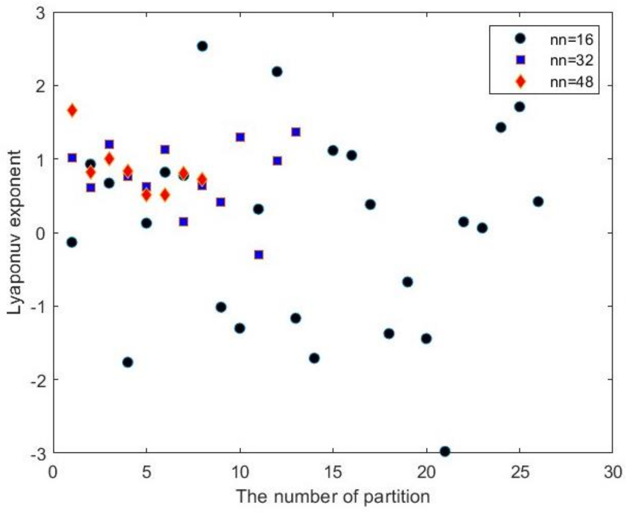


FIGURE 11. Lyapunov exponent versus number of partitions for the 468-landmark 3D facial image dataset.

Compared to deep learning (DL) algorithms, the proposed methodology is computationally efficient (faster and requires limited memory), does not require larger data sets or higher computational power to identify facial emotions. PRF classifiers are highly accurate in predicting emotions, but the proposed approach has a few limitations.

- One limitation is that the present experimental framework utilized a database [25] developed using virtual markers in real-time emotion recognition tasks. Most facial emotion recognition databases in the literature consist of static facial images [27], [28] or video-based emotion recognition. To the best of our knowledge, currently, there are no databases for real-time facial emotion recognition using virtual markers. Therefore, a comparison between the efficiency of the proposed method on different datasets was not made. To show the efficiency and robustness of the proposed method, a real-time 3D FER dataset [67] was also considered. However, benchmarking the results of this experiment with other open-source databases would be extremely challenging.
- Another limitation is that this study investigated the chaotic behavior of facial features by using three nonlinear measures: LE, AE, and CD. Other nonlinear features such as Fractal Dimension (FD), Recurrence Quantification Analysis (RQA), and others could be analyzed to assess the PRF classifier's performance in recognizing facial emotions with the three proposed features.

In light of this, future work should focus on the following: (a) extending PRF to other practical datasets, (b) identifying the most suitable algorithm for data partitioning across a variety of datasets, (c) developing a theoretical approach to determinant’s optimal value, and (d) analyzing statistical characteristics of classification accuracy.

V. CONCLUSION

The rapid development of digital technology and human-computer interaction has made facial emotion recognition an intriguing topic from a theoretical and practical point of view. This paper developed a novel variant of the random forest method, the partitioned random forest method, to improve facial emotion recognition. The main idea is to divide large datasets into smaller disjoint datasets according to a predetermined order, achieved by applying the classification method to each dataset. Although existing classification accuracies are promising, over 90% of subsets will significantly improve classification accuracy using the presented method. Results reveal that Lyapunov exponents and the number of elements are significant factors in classification accuracy. Subsets with large positive Lyapunov exponents and smaller elements demonstrated higher classification accuracy, indicating that an increased number of samples does not necessarily mean greater accuracy in random forest classification. It is also evident that using the norm-1 of each element for partitioning leads to a significant increase in classification accuracy. Future work could involve applying the proposed method to feature selection in Random Forest (RF).

ACKNOWLEDGEMENT

The open access publication cost is supported by Qatar National Library.

COMPETING INTERESTS

The authors declare no competing interests.

DATA AVAILABILITY

Data is available by request to the corresponding author MM.

REFERENCES

[1] A. Hassouneh, A. M. Mutawa, and M. Murugappan, "Development of a Real-Time Emotion Recognition System Using Facial Expressions and EEG based on machine learning and deep neural network methods," Informatics in Medicine Unlocked, vol. 20, p. 100372, 2020, doi: 10.1016/j.imu.2020.100372.

[2] E. Batbaatar, M. Li, and K. H. Ryu, "Semantic-Emotion Neural Network for Emotion Recognition From Text," IEEE Access, vol. 7, pp. 111866–111878, 2019, doi: 10.1109/ACCESS.2019.2934529.

[3] B. Schuller, G. Rigoll, and M. Lang, "Hidden Markov model-based speech emotion recognition," in 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03)., Hong Kong, China: IEEE, 2003, p. II-1–4. doi: 10.1109/ICASSP.2003.1202279.

[4] R. A. Khalil, E. Jones, M. I. Babar, T. Jan, M. H. Zafar, and T. Alhussain, "Speech Emotion Recognition Using Deep Learning Techniques: A Review," IEEE Access, vol. 7, pp. 117327–117345, 2019, doi: 10.1109/ACCESS.2019.2936124.

[5] A. Koduru, H. B. Valiveti, and A. K. Budati, "Feature extraction algorithms to improve the speech emotion recognition rate," Int J Speech Technol, vol. 23, no. 1, pp. 45–55, Mar. 2020, doi: 10.1007/s10772-020-09672-4.

[6] S. Zepf, J. Hernandez, A. Schmitt, W. Minker, and R. W. Picard, "Driver Emotion Recognition for Intelligent Vehicles: A Survey," ACM Comput. Surv., vol. 53, no. 3, pp. 1–30, May 2021, doi: 10.1145/3388790.

[7] M. Maithri et al., "Automated emotion recognition: Current trends and future perspectives," Computer Methods and Programs in Biomedicine, vol. 215, p. 106646, Mar. 2022, doi: 10.1016/j.cmpb.2022.106646.

[8] B. Ko, "A Brief Review of Facial Emotion Recognition Based on Visual Information," Sensors, vol. 18, no. 2, p. 401, Jan. 2018, doi: 10.3390/s18020401.

[9] S. C. Leong, Y. M. Tang, C. H. Lai, and C. K. M. Lee, "Facial expression and body gesture emotion recognition: A systematic review on the use of visual data in affective computing," Computer Science Review, vol. 48, p. 100545, May 2023, doi: 10.1016/j.cosrev.2023.100545.

[10] S. Nathani, "A Comparative Study of Transfer Learning for Emotion Recognition using CNN and Modified VGG16 Models," Jul. 19, 2024, arXiv: arXiv:2407.14576. doi: 10.48550/arXiv.2407.14576.

[11] K. Zaman et al., "A novel driver emotion recognition system based on deep ensemble classification," Complex Intell. Syst., vol. 9, no. 6, pp. 6927–6952, Dec. 2023, doi: 10.1007/s40747-023-01100-9.

[12] A. P. Fard, M. M. Hosseini, T. D. Sweeny, and M. H. Mahoor, "AffectNet+: A Database for Enhancing Facial Expression Recognition with Soft-Labels," Oct. 29, 2024, arXiv: arXiv:2410.22506. doi: 10.48550/arXiv.2410.22506.

[13] P. Naga, S. D. Marri, and R. Borreo, "Facial emotion recognition methods, datasets and technologies: A literature survey," Materials Today: Proceedings, vol. 80, pp. 2824–2828, 2023, doi: 10.1016/j.matpr.2021.07.046.

[14] S. Argaud, M. Vérin, P. Sauleau, and D. Grandjean, "Facial emotion recognition in Parkinson's disease: A review and new hypotheses," Movement Disorders, vol. 33, no. 4, pp. 554–567, Apr. 2018, doi: 10.1002/mds.27305.

[15] K. Slimani, M. Kas, Y. El Merabet, R. Messoussi, and Y. Ruichek, "Facial emotion recognition: A comparative analysis using 22 LBP variants," in Proceedings of the 2nd Mediterranean Conference on Pattern Recognition and Artificial Intelligence, Rabat Morocco: ACM, Mar. 2018, pp. 88–94. doi: 10.1145/3177148.3180092.

[16] D. V. Sang, L. T. B. Cuong, and P. T. Ha, "Discriminative Deep Feature Learning for Facial Emotion Recognition," in 2018 1st International Conference on Multimedia Analysis and Pattern Recognition (MAPR), Ho Chi Minh City: IEEE, Apr. 2018, pp. 1–6. doi: 10.1109/MAPR.2018.8337514.

[17] M. A. H. Akhand, S. Roy, N. Siddique, M. A. S. Kamal, and T. Shimamura, "Facial Emotion Recognition Using Transfer Learning in the Deep CNN," Electronics, vol. 10, no. 9, p. 1036, Apr. 2021, doi: 10.3390/electronics10091036.

[18] N. Rathour et al., "IoMT Based Facial Emotion Recognition System Using Deep Convolution Neural Networks," Electronics, vol. 10, no. 11, p. 1289, May 2021, doi: 10.3390/electronics10111289.

[19] A. Alreshidi and M. Ullah, "Facial Emotion Recognition Using Hybrid Features," Informatics, vol. 7, no. 1, p. 6, Feb. 2020, doi: 10.3390/informatics7010006.

[20] F. Z. Canal et al., "A survey on facial emotion recognition techniques: A state-of-the-art literature review," Information Sciences, vol. 582, pp. 593–617, Jan. 2022, doi: 10.1016/j.ins.2021.10.005.

[21] M. A. Haghpanah, E. Saeedizade, M. T. Masouleh, and A. Kalhor, "Real-Time Facial Expression Recognition using Facial Landmarks and Neural Networks," in 2022 International Conference on Machine Vision and Image Processing (MVIP), Ahvaz, Iran, Islamic Republic of: IEEE, Feb. 2022, pp. 1–7. doi: 10.1109/MVIP53647.2022.9738754.

[22] D. Lakshmi and R. Ponnusamy, "Facial emotion recognition using modified HOG and LBP features with deep stacked autoencoders," Microprocessors and Microsystems, vol. 82, p. 103834, Apr. 2021, doi: 10.1016/j.micpro.2021.103834.
[23] L. Graumann et al., "Facial emotion recognition in borderline patients is unaffected by acute psychosocial stress," Journal of Psychiatric Research, vol. 132, pp. 131–135, Jan. 2021, doi: 10.1016/j.jpsychires.2020.10.007.
[24] N. Jain, S. Kumar, A. Kumar, P. Shamsolmoali, and M. Zareapoor, "Hybrid deep neural networks for face emotion recognition," Pattern Recognition Letters, vol. 115, pp. 101–106, Nov. 2018, doi: 10.1016/j.patrec.2018.04.010.
[25] M. Murugappan and A. Mutawa, "Facial geometric feature extraction based emotional expression classification using machine learning algorithms," PLoS ONE, vol. 16, no. 2, p. e0247131, Feb. 2021, doi: 10.1371/journal.pone.0247131.
[26] Y. Khaireddin and Z. Chen, "Facial Emotion Recognition: State of the Art Performance on FER2013," May 08, 2021, arXiv: arXiv:2105.03588. Accessed: Mar. 21, 2024. [Online]. Available: http://arxiv.org/abs/2105.03588
[27] M. Dirik, O. Castillo, and A. F. Kocamaz, "Emotion Recognition Based on Interval Type-2 Fuzzy Logic from Facial Expression," Journal of Soft Computing and Artificial Intelligence, vol. 1, pp. 1–17, 2020.
[28] M. Dirik, "Optimized Anfis Model with Hybrid Metaheuristic Algorithms for Facial Emotion Recognition," Int. J. Fuzzy Syst., vol. 25, no. 2, pp. 485–496, Mar. 2023, doi: 10.1007/s40815-022-01402-z.
[29] W. Mellouk and W. Handouzi, "Facial emotion recognition using deep learning: review and insights," Procedia Computer Science, vol. 175, pp. 689–694, 2020, doi: 10.1016/j.procs.2020.07.101.
[30] Cuicui Zhang, T. Matsuyama, and Xuefeng Liang, "Small Sample Size Face Recognition using Random Quad-Tree based Ensemble Algorithm," in 5th International Conference on Imaging for Crime Detection and Prevention (ICDP 2013), London, UK: Institution of Engineering and Technology, 2013, p. 2.02-2.02. doi: 10.1049/ic.2013.0270.
[31] Y. Li, "Face Detection Algorithm Based on Double-Channel CNN with Occlusion Perceptron," Computational Intelligence and Neuroscience, vol. 2022, pp. 1–10, Jan. 2022, doi: 10.1155/2022/3705581.
[32] A. Althnian et al., "Impact of Dataset Size on Classification Performance: An Empirical Evaluation in the Medical Domain," Applied Sciences, vol. 11, no. 2, p. 796, Jan. 2021, doi: 10.3390/app11020796.
[33] T. Oyedare and J.-M. J. Park, "Estimating the Required Training Dataset Size for Transmitter Classification Using Deep Learning," in 2019 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN), Newark, NJ, USA: IEEE, Nov. 2019, pp. 1–10. doi: 10.1109/DySPAN.2019.8935823.
[34] C. Chu, A.-L. Hsu, K.-H. Chou, P. Bandettini, and C. Lin, "Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images," NeuroImage, vol. 60, no. 1, pp. 59–70, Mar. 2012, doi: 10.1016/j.neuroimage.2011.11.066.
[35] A. Bailly et al., "Effects of dataset size and interactions on the prediction performance of logistic regression and deep learning models," Computer Methods and Programs in Biomedicine, vol. 213, p. 106504, Jan. 2022, doi: 10.1016/j.cmpb.2021.106504.
[36] H. Gong, Y. Sun, X. Shu, and B. Huang, "Use of random forests regression for predicting IRI of asphalt pavements," Construction and Building Materials, vol. 189, pp. 890–897, Nov. 2018, doi: 10.1016/j.conbuildmat.2018.09.017.
[37] M. I. N. P. Munasinghe, "Facial Expression Recognition Using Facial Landmarks and Random Forest Classifier," in 2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS), Singapore: IEEE, Jun. 2018, pp. 423–427. doi: 10.1109/ICIS.2018.8466510.
[38] K. Tanaka, T. Kinkyo, and S. Hamori, "Random forests-based early warning system for bank failures," Economics Letters, vol. 148, pp. 118–121, Nov. 2016, doi: 10.1016/j.econlet.2016.09.024.
[39] S. Wang, C. Aggarwal, and H. Liu, "Random-Forest-Inspired Neural Networks," ACM Trans. Intell. Syst. Technol., vol. 9, no. 6, pp. 1–25, Nov. 2018, doi: 10.1145/3232230.
[40] M. Jeong and B. C. Ko, "Driver's Facial Expression Recognition in Real-Time for Safe Driving," Sensors, vol. 18, no. 12, p. 4270, Dec. 2018, doi: 10.3390/s18124270.
[41] O. Pauly, "Random Forests for Medical Applications," TECHNISCHE UNIVERSITÄT MÜNCHEN, 2012.
[42] Y. Wang, Y. Li, Y. Song, and X. Rong, "Facial Expression Recognition Based on Random Forest and Convolutional Neural Network," Information, vol. 10, no. 12, p. 375, Nov. 2019, doi: 10.3390/info10120375.
[43] A. Ziegler and I. R. König, "Mining data with random forests: current options for real-world applications," WIREs Data Min & Knowl, vol. 4, no. 1, pp. 55–63, Jan. 2014, doi: 10.1002/widm.1114.
[44] S. Alsubai, "Emotion Detection Using Deep Normalized Attention-Based Neural Network and Modified-Random Forest," Sensors, vol. 23, no. 1, p. 225, Dec. 2022, doi: 10.3390/s23010225.
[45] A. Dapogny and K. Bailly, "Investigating Deep Neural Forests for Facial Expression Recognition," in 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), Xi'an: IEEE,

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3560362

May 2018, pp. 629–633. doi: 10.1109/FG.2018.00099.

[46] S. Gharsalli, B. Emile, H. Laurent, X. Desquesnes, and D. Vivet, "Random forest-based feature selection for emotion recognition," in 2015 International Conference on Image Processing Theory, Tools and Applications (IPTA), Orleans, France: IEEE, Nov. 2015, pp. 268–272. doi: 10.1109/IPTA.2015.7367144.

[47] Z. Sun, G. Wang, P. Li, H. Wang, M. Zhang, and X. Liang, "An improved random forest based on the classification accuracy and correlation measurement of decision trees," Expert Systems with Applications, vol. 237, p. 121549, Mar. 2024, doi: 10.1016/j.eswa.2023.121549.

[48] J. Chen et al., "A Parallel Random Forest Algorithm for Big Data in a Spark Cloud Computing Environment," IEEE Trans. Parallel Distrib. Syst., vol. 28, no. 4, pp. 919–933, Apr. 2017, doi: 10.1109/TPDS.2016.2603511.

[49] L. Yin, K. Chen, Z. Jiang, and X. Xu, "A Fast Parallel Random Forest Algorithm Based on Spark," Applied Sciences, vol. 13, no. 10, p. 6121, May 2023, doi: 10.3390/app13106121.

[50] M. Shahhosseini and G. Hu, "Improved Weighted Random Forest for Classification Problems," in Progress in Intelligent Decision Science, vol. 1301, T. Allahviranloo, S. Salahshour, and N. Arica, Eds., in Advances in Intelligent Systems and Computing, vol. 1301. , Cham: Springer International Publishing, 2021, pp. 42–56. doi: 10.1007/978-3-030-66501-2_4.

[51] K. Gubbala, M. N. Kumar, and A. M. Sowjanya, "AdaBoost based Random forest model for Emotion classification of Facial images," MethodsX, vol. 11, p. 102422, Dec. 2023, doi: 10.1016/j.mex.2023.102422.

[52] V. Y. Kulkarni and P. K. Sinha, "Efficient Learning of Random Forest Classifier using Disjoint Partitioning Approach," presented at the Proceedings of the World Congress on Engineering, 2013.

[53] Y. Liu and Z. Ge, "Weighted random forests for fault classification in industrial processes with hierarchical clustering model selection," Journal of Process Control, vol. 64, pp. 62–70, Apr. 2018, doi: 10.1016/j.jprocont.2018.02.005.

[54] K. Shah, H. Patel, D. Sanghvi, and M. Shah, "A Comparative Analysis of Logistic Regression, Random Forest and KNN Models for the Text Classification," Augment Hum Res, vol. 5, no. 1, p. 12, Dec. 2020, doi: 10.1007/s41133-020-00032-0.

[55] S. Han, B. D. Williamson, and Y. Fong, "Improving random forest predictions in small datasets from two-phase sampling designs," BMC Med Inform Decis Mak, vol. 21, no. 1, p. 322, Dec. 2021, doi: 10.1186/s12911-021-01688-3.

[56] V. F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez, "An assessment of the effectiveness of a random forest classifier for land-cover classification," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 67, pp. 93–104, Jan. 2012, doi: 10.1016/j.isprsjprs.2011.11.002.

[57] A. Rácz, D. Bajusz, and K. Héberger, "Effect of Dataset Size and Train/Test Split Ratios in QSAR/QSPR Multiclass Classification," Molecules, vol. 26, no. 4, p. 1111, Feb. 2021, doi: 10.3390/molecules26041111.

[58] C. Catal and B. Diri, "Investigating the effect of dataset size, metrics sets, and feature selection techniques on software fault prediction problem," Information Sciences, vol. 179, no. 8, pp. 1040–1058, Mar. 2009, doi: 10.1016/j.ins.2008.12.001.

[59] F. E. Nowruzi, P. Kapoor, D. Kolhatkar, F. A. Hassanat, R. Laganiere, and J. Rebut, "How much real data do we actually need: Analyzing object detection performance using synthetic and real data," Jul. 16, 2019, arXiv: arXiv:1907.07061. Accessed: Mar. 21, 2024. [Online]. Available: http://arxiv.org/abs/1907.07061

[60] P. Gupta, S. Bhaskarpandit, and M. Gupta, "Similarity Learning based Few Shot Learning for ECG Time Series Classification," in 2021 Digital Image Computing: Techniques and Applications (DICTA), Gold Coast, Australia: IEEE, Nov. 2021, pp. 1–8. doi: 10.1109/DICTA52665.2021.9647357.

[61] S. Shahinfar, P. Meek, and G. Falzon, "'How many images do I need?' Understanding how sample size per class affects deep learning model performance metrics for balanced designs in autonomous wildlife monitoring," Ecological Informatics, vol. 57, p. 101085, May 2020, doi: 10.1016/j.ecoinf.2020.101085.

[62] M. Goudjil, M. Koudil, M. Bedda, and N. Ghoggali, "A Novel Active Learning Method Using SVM for Text Classification," Int. J. Autom. Comput., vol. 15, no. 3, pp. 290–298, Jun. 2018, doi: 10.1007/s11633-015-0912-z.

[63] H. Wang, S. Ge, Z. Lipton, and E. P. Xing, "Learning Robust Global Representations by Penalizing Local Predictive Power," in Advances in Neural Information Processing Systems, 2019.

[64] A. Velichko, M. Belyaev, Y. Izotov, M. Murugappan, and H. Heidari, "Neural Network Entropy (NNetEn): EEG Signals and Chaotic Time Series Separation by Entropy Features, Python Package for NNetEn Calculation".

[65] J. Fan, S. Upadhye, and A. Worster, "Understanding receiver operating characteristic (ROC) curves," CJEM, vol. 8, no. 01, pp. 19–20, Jan. 2006, doi: 10.1017/S1481803500013336.

[66] H. Heidari and A. A. Velichko, "An improved LogNNet classifier for IoT applications," J. Phys.: Conf. Ser., vol. 2094, no. 3, p. 032015, Nov. 2021, doi: 10.1088/1742-6596/2094/3/032015.

[67] M. J. Ashraf, "Emotion_Recognition_Mediapipe Public."

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3560362

https://github.com/JafirDon/Emotion_Recognition_Mediapipe, 2021.

HANIF HEIDARI is an assistant professor in the Department of Applied Mathematics, Damghan University, Damghan, Iran. His research interests are quantum machine learning, time series classification and prediction, grey system theory, and theoretical aspects of metaheuristic optimization methods. With over twelve years of experience, his foundations have been published in high-quality interdisciplinary scientific journals. He has also given more than 20 talks at international conferences and supervised 22 graduate students in applied mathematics and computer science. Hanif enjoys international scientific collaboration for solving complex practical problems using mathematical methods.

M. MURUGAPPAN has been working at Kuwait College of Science and Technology (KCST), Kuwait as a Full Professor in Electronics, Department of Electronics and Communication Engineering since 2016. He is also serving as a Visiting Professor at the School of Engineering at Vels Institute of Science, Technology, and Advanced Studies in India and an International Visiting Fellow at the Center of Excellence in Unmanned Aerial Systems at Universiti Malaysia Perlis in Malaysia. In 2006, he graduated from Anna University, India with an M.E. degree in Applied Electronics. He received his Ph.D. from Universiti Malaysia Perlis, Malaysia in 2010 for his contribution to the field of Mechatronic Engineering. Between 2010 and 2016, he worked as a Senior Lecturer at the School of Mechatronics Engineering, Universiti Malaysia Perlis, Malaysia. In a study by Stanford University, he was recently ranked among the top 2 percent of scientists working in experimental psychology and artificial intelligence for three consecutive years (2020-2022). To date, his Google Scholar citations reach 8850+ with an H-Index of 47. His research into affective computing has received more than 750K in grants from Malaysia, Kuwait, and the UK. His publications include more than 170 peer-reviewed conference proceedings papers, journal articles, and book chapters. Several of his journal articles have been recognized as best papers, best papers of the fiscal year, etc. Prof. Murugappan is a member of the editorial boards of PlosOne, Human Centric Information Sciences, PEERJ Computer Science, Journal of Medical Imaging and Health Informatics, and International Journal of Cognitive Informatics. In addition to being the Chair of the IEEE Kuwait Section's Educational Activities Committee, Kuwait. He is interested in affective computing, affective neuroscience, neuromarketing, and medical image processing.

JAVEED SHAIKH-MOHAMMED received the B.Tech. degree in electrical and electronics engineering from Sri Venkateswara University, India, and the M.S. degree in electrical engineering and the Ph.D. degree in engineering from Louisiana Tech University, USA. He is currently an Associate Professor in the Biomedical Technology Department at Prince Sattam bin Abdulaziz University, Saudi Arabia. He previously worked at IIT-M (India), UniSZA (Malaysia), VIT (India), KSU (Saudi Arabia), UW-Madison (USA), and UIC-Chicago (USA). His research interests include BioMEMS, nanotechnology, microfluidics, assistive technologies, and artificial intelligence.

MUHAMMAD E. H. CHOWDHURY received his Ph.D. degree from the University of Nottingham, U.K., in 2014. He worked as a Postdoctoral Research Fellow at the Sir Peter Mansfield Imaging Centre, University of Nottingham. He is currently working as an Assistant Professor with the Department of Electrical Engineering, at Qatar University. He has filed several patents and published more than 180 peer-reviewed journal articles, 30+ conference papers, and several book chapters. His current research interests include biomedical instrumentation, signal processing, wearable sensors, medical image analysis, machine learning and computer vision, embedded system design, and simultaneous EEG/fMRI. He is currently running NPRP, UREP, and HSREP grants from the Qatar National Research Fund (QNRF) and internal grants (IRCC and HIG) from Qatar University along with academic projects from HBKU and HMC. He is a Senior Member of IEEE, and a member of British Radiology, ISMRM, and HBM. He serves as a Guest Editor for Polymers, an Associate Editor for IEEE Access, and a Topic Editor and Review Editor for Frontiers in Neuroscience. He has recently won the COVID-19 Dataset Award, AHS Award from HMC, and National AI Competition awards for contributing to the fight against COVID-19. His team is the gold medalist at the 13th International Invention Fair in the Middle East (IIFME). He has been listed among the Top 2% of scientists in the World List, published by Stanford University.