IEEE Access
SPECIAL SECTION ON ADVANCES IN MACHINE LEARNING AND COGNITIVE COMPUTING FOR INDUSTRY APPLICATIONS

Received June 27, 2020, accepted July 7, 2020, date of publication July 17, 2020, date of current version August 3, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3009908

Artificial Neural Network Based Ensemble Approach for Multicultural Facial Expressions Analysis

GHULAM ALI[1], AMJAD ALI[2], FARMAN ALI[3], UMAR DRAZ[2,4], FIAZ MAJEED[5], SANA YASIN[2], TARIQ ALI[6], AND NOMAN HAIDER[7]

[1]Department of Computer Science, University of Okara, Okara 56300, Pakistan
[2]Department of Computer Science, COMSATS University Islamabad–Lahore, Lahore 54000, Pakistan
[3]Department of Software, Sejong University, Seoul 05006, South Korea
[4]Department of Computer Science, University of Sahiwal, Sahiwal 57000, Pakistan
[5]Department of Software Engineering, University of Gujrat, Gujrat 50700, Pakistan
[6]Department of Electrical Engineering, College of Engineering, Najran University, Najran 61441, Saudi Arabia
[7]College of Engineering and Science, Victoria University, Sydney, NSW 2000, Australia

Corresponding authors: Amjad Ali (amjad.ali@cuilahore.edu.pk) and Farman Ali (farmankanju@sejong.ac.kr)

This work was supported in part by the Sejong University through its faculty research program.

ABSTRACT Facial expressions convey exhaustive information about human emotions and the most interactive way of social collaborations, despite differences in ethnicity, culture, and geography. Due to cultural differences, the variations in facial structure, facial appearance, and facial expression representation are the main challenges to the facial expression recognition system. These variations necessitate the need for multicultural facial expression analysis. This study presents several computational algorithms to handle these variations to get high expression recognition accuracy. We propose an artificial neural network-based ensemble classifier for multicultural facial expression analysis. The facial images from the Japanese female facial expression database, Taiwanese facial expression image database, and RadBoud faces database are combined to form a multi-culture facial expression dataset. The participants in the multicultural dataset originate from four ethnic regions including Japanese, Taiwanese, Caucasians, and Moroccans. Local binary pattern, uniform local binary pattern, and principal component analysis are applied for facial feature representation. Experimental results prove that facial expressions are innate and universal across all cultures with minor variations.

INDEX TERMS Facial expression, multicultural, ensemble, artificial neural network.

I. INTRODUCTION

The cultural variation in facial expression representation raises the issue of the universality of facial expressions. In psychological and social science literature, the universality of human emotions in the appearance of facial expressions has remained one of the highest standing debates. In the late 1960s and early 1970s, researchers found evidence about the universality of facial expressions [1], [2]. Ekman identified basic emotions (i.e., happiness, anger, surprise, sadness, fear, and disgust) which are common for all human across different cultures [5]. In our research, we studied five expressions, such as anger, happiness, surprise, sadness, and fear for cross-cultural facial expression categorization. The facial expression has always been a challenging problem in the field of machine learning, image processing, and sample recognition. It plays an important role in representing the internal feeling state in a computer-assisted human interaction environment.

Due to such importance, a lot of effort has been devoted to developing reliable facial expression classification techniques. The smart meeting, visual surveillance, social robotics, and video conferencing are some interesting real-world applications which require cross-cultural facial expression recognition, many machine learning techniques [4]–[7] have been developing for culture-specific facial expression recognition, but no effective technique has been developed to recognize the multicultural facial expressions efficiently [8], [9]. The reason behind is that the facial structure has too many variations due to difference in

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

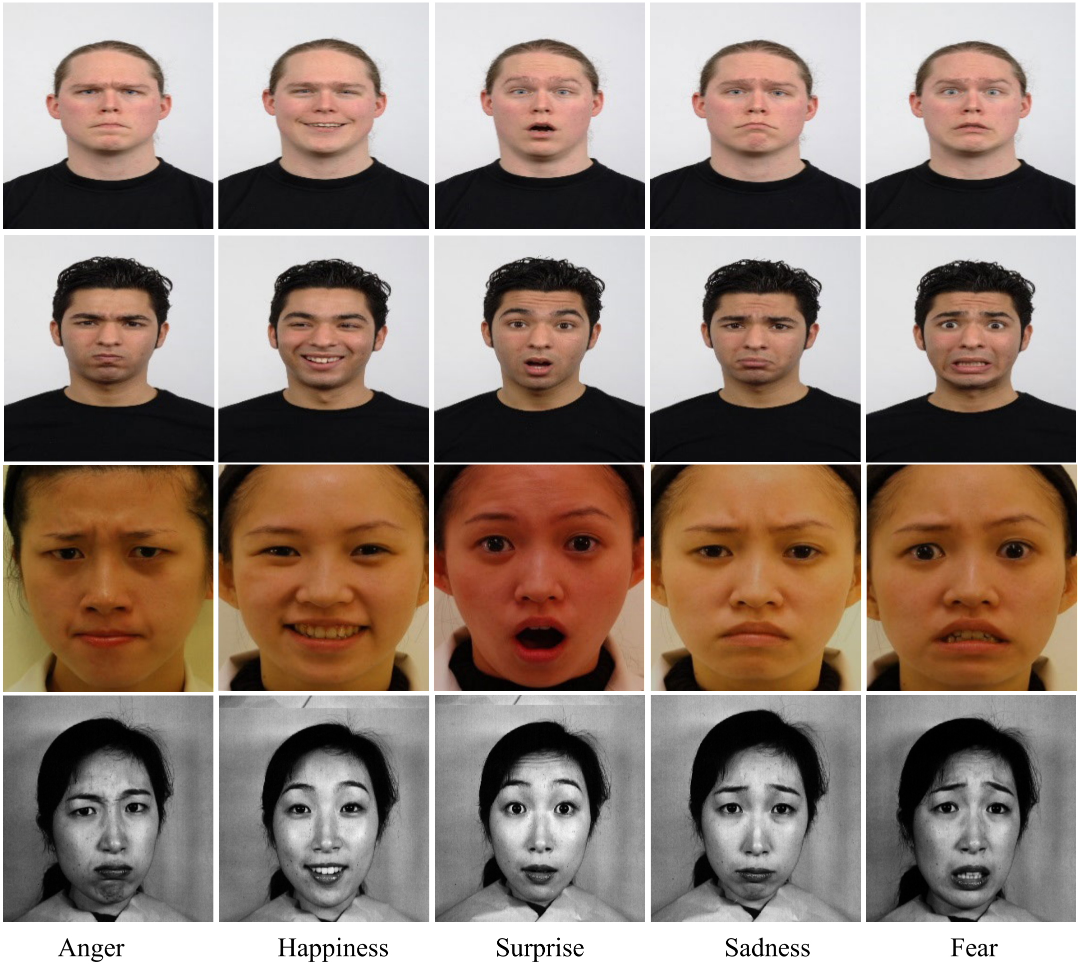


FIGURE 1. **Multicultural facial structure and facial expression representation variations from (row 1) Moroccan to (row 2) Caucasian, (row 3) Taiwanese, and (row 4) Japanese.**

culture, color, and the cultural state as shown in Fig. 1. Due to this diversity, the existing facial expression classification approaches focus only on the culturally specific facial expression recognition. Therefore, more research efforts are compulsory for multicultural facial expression recognition techniques. This research focuses on the development of a neural network-based ensemble technique to perfectly identify the cross-cultural facial expression representation variations. The identifiers should be talented to learn the variations among Moroccan, Caucasian, cultures with the Japanese and Taiwanese subjects.

We have developed a novel ensemble approach called artificial neural network ensemble collection (ANNE). The developed ensemble model consists of three-level of classifiers, base level, meta-level, and predictor. A collection of binary neural networks (BNNs) is experienced to form an ANNE collection. The outputs of BNNs in an ANNE are combined using a probability distribution, where each ANNE represents the possible facial expression. Thus, the use of a binary scheme allows each ANNE to admit or negligible the survival of an expression. We believe that the proposed system functions irrespective of constraints like culture, color, ethnic and geographical region.

II. CONTRIBUTIONS

In this research, an ensemble technique called ensemble collections is applied to those peoples which belong to different cultures, this type of dataset identifies the different multicultural facial expressions by making the following contributions:

- We developed the multi-culture facial expression database by combining the three different facial expression databases: Japanese female facial expression (JAFFE), Taiwanese facial expression image database (TFEID), and RadBoud face database (RaFD). The participants of the multi-culture database originate from following cultural, ethnic, and geographic regions; Moroccan, Caucasian, Japanese, and Taiwanese.
- We applied a two-level feature extraction approach, in which the LBP, the PCA, and the ULBP filters were used as the preprocessing stage for feature extraction, representing the facial features with many dimensions.

Furthermore, the dimensionality of the aspect vector was decreased using the PCA algorithm.

- A novel ensemble approach called ensemble collections is proposed to learn the cross-cultural facial expression variations. The proposed ensemble classifier consists of three levels of classifiers: base-level, meta-level, and predictor.
- We developed a novel meta-learning technique, which is the fusion of NB classifier and Bernoulli distribution for well-organized cross-cultural facial expression which is classified by learning the presence of facial expression from a vast collection of ANNEs.

III. LITERATURE REVIEW

A comprehensive review of human emotion perception techniques and issues up to 2009 are presented in [10]. The authors discussed the challenges and data availability issues also they extend to other types of media that can be used to recognize the spontaneous human emotions. Lajevardi *et al.* in [11] represented a comparative study of different facial expression categorization methods and feature extraction techniques. Donato *et al.* in [12] presented a comprehensive comparison of facial feature representation and dimension reduction techniques. Zhao *et al.* in [13] proposed scheme for facial region detection and facial expression identification. Recently, Torre and Cohn [14] reviewed the state-of-the-art facial expression recognition methods. The authors analyzed essential approaches and current efforts including feature extraction, face tracking, and facial feature representation.

The main objective of the ensemble technique is to combine a set of classifiers, where each classifier classifies the same original class. The decisions of classifiers are combined to obtain the best global classifier, with more accurate and reliable decisions instead of using only one classifier. The experimental studies in machine learning show that combining the decisions of various classifiers reduces the simplification error. The most popular techniques for construct ensembles are bagging, boosting, and stacking. The objective of constructing an ensemble of classifiers is not only the expression recognition error minimization but also to optimize the number of classifiers in the ensemble classifier. The selection of an optimal number of classifiers from a pool of classifiers while maintaining classifiers' accuracy is an optimization problem [15]. In this regard, many optimization techniques have been used for ensemble construction. Kuncheva *et al.* in [16] employed a genetic algorithm to select the best combination of classifiers from three different sets of classifiers and achieve capable exactness as compare to the single classifier. Similarly, Kortemeyer *et al.* in [17] introduced a genetic algorithm-based ensemble approach to enhance classifier efficiency and to achieve better classification accuracy. In [18] a heritable algorithm-based ensemble design approach is urbanized where the training data re-sampling and weak classifiers evolution maximizes the classification accuracy. It proved that the optimization-based ensemble approach gives better classification accuracy as compared to AdaBoost.

Significantly, neural network-based ensemble techniques are being used to enhance the expression recognition accuracy. Chen *et al.* [19] proposed an ANN-based ensemble technique for facial expression recognition. The bagging ensemble manufacturing approach was applied to the instructor the ANN classifiers. Wang and Xiao [4] proposed a neural network ensemble technique using the Md-Adaboost ensemble construction approach to combine the results of a pool of neural networks. The artificial neural networks were trained using different combinations of feature vectors to enhance expression recognition accuracy and classifiers' stability. Windeatt [20] described the various coding and decoding techniques, which are used in ECOC ensemble construction. The ECOC was modified to tune the parameters of artificial neural networks.

Extreme learning machine (ELM) is a new extension of neural network-based learning techniques, which provide the greatest generalization performance at exceptionally fast learning speed [21]. On the other hand, the authors were failing to present the criteria of judgment for the selection of parameters such as the number of secreted layers and numbers in each hidden layer. Recently, an extreme learning machine-based ensemble approach has been developed for facial expression recognition [6]. A poll of ELMs was trained using a bagging ensemble construction approach. The experiments were performed on JAFFE and CK+ datasets to evaluate the performance of the ensemble classifier. More recently, Wang and Xiao [22] proposed a novel fuzzy integral technique to integrate the facial expression predictions of multiple neural networks. Extensive experiments were performed on the JAFFE dataset to evaluate the validity of the fuzzy integral ensemble approach. Ali *et al.* [7] introduced a neural network based boosted ensemble approach for facial expression recognition. A set of neural network classifiers trained using the histogram of oriented gradients by varying the neural network architecture. The final prediction was made based on the weighted average of all neural networks' decisions. Kim *et al.* [23] developed a deep Convolutional neural network-based ensemble committee for facial expression recognition. Multiple deep CNNs were trained to vary the network architecture, input normalization, and network initial weights. To improve the expression recognition accuracy, the ensemble committee was organized in a hierarchical structure with exponentially weighted decision fusion. The expression recognition accuracy was very low on three facial expression databases. The major drawback of this technique is the use of exponentially weighted decision fusion instead of the Meta-learning approach.

Meta-learning in ensemble models, learn how the classifier can increase its efficiency through experience. The objective is to understand how the classifier itself can become flexible according to classification tasks. In meta-learning, the meta-classifier learns from the errors of base-level classifiers to accurately predict the presence of an expression. In the training of the meta-classifier, the major problem is the formation of the training set to train the classifier. Chen and Wong [24] proposed an ant colony optimization approach with a stacking ensemble classifier. They performed some preliminary experiments to compare the proposed approach with many existing ensemble techniques such as bagging, boosting, and ECOC. Ghimire and Lee [25] employed time wrapping similarity distance as a weak classifier to select the discriminative feature vectors and support vector machines to predict the presence of an expression.

As one of the recent deep learning-based facial expression recognition approaches, Villanueva *et al.* in [26] developed a real-time facial expression recognition system to recognize the happiness and sadness expressions. The proposed approach is robust to variations in facial expression representation, ethnicity, culture, age, geography, and pose. Yu *et al.* in [27] proposed a deep learning technique to mitigate the problem of inter-expression resemblance and intra-expression variation. First extracted the region of interest from facial images then applied deep metric learning and partial image method to recognize the facial expression more accurately. Similarly, Shi *et al.* [28] applied the FCM on the convolution layer of CNN to achieve high expression recognition accuracy with low execution time. The authors in [45]–[48] proposed multimedia transmission schemes for cognitive radio networks.

The evidence about the universality of facial expressions is presented in [3]. It demonstrates that each expression has unique features (gesture, structure, and internal emotional state) as well as some common characteristics with other expressions (rapid onset, short duration, unbidden occurrence, automatic appraisal, and coherence among responses). The effects of these unique and common characteristics on facial expression representation vary from culture to culture. Russell [29] defined that emotions of anger, fear, and contempt are universal that naturally evolve through the culture. In [30] the authors examined similarities and differences in cross-cultural expression recognition. Find out that facial expressions across different cultures are universal with subtle differences.

Until early 2009, the issue of cross-cultural facial expression recognition was comparatively ignored. Because, the use of universal facial expression dataset is rare, and comparatively less attention has been given to the problem of universal facial expression recognition. Matsumoto *et al.* [31] performed cross-culture emotions analysis on spontaneously produced facial expressions by American, Japanese, and British students. The signal clarity of each expression was similar across cultures. Dailey *et al.* [32] developed a machine learning approach to recognize cross-culture facial expression recognition. The authors concluded that minute variations in multicultural facial images can affect classifier performance. Da Silva and Pedrini [8] performed a similar type of experiment. The authors suggested that six basic facial expressions are natural and universal with minor variations. The authors developed a multicultural dataset by combining the four different benchmark databases. The proposed technique was unable to define a suitable combination of features and classifiers to recognition the multicultural facial expressions efficiently. Similarly, Zia and Jaffar [9] proposed a novel incremental learning technique to efficiently recognize multicultural facial expression. The proposed incremental learning technique enables the predictor to learn the cultural variation incrementally. The classifiers were trained incrementally on four different facial expression databases. The performance of trained predictors was evaluated on the TEST facial expression database, which was developed by collecting different facial images of female's actresses from different sources. The proposed techniques performed significantly as compared to state-of-the-art techniques. However, the multicultural facial expression recognition techniques presented in [8], [9], [32] motivates toward the development of a diverse multicultural facial expression database as well as more sophisticated prediction techniques that can efficiently recognize the multicultural facial expressions.

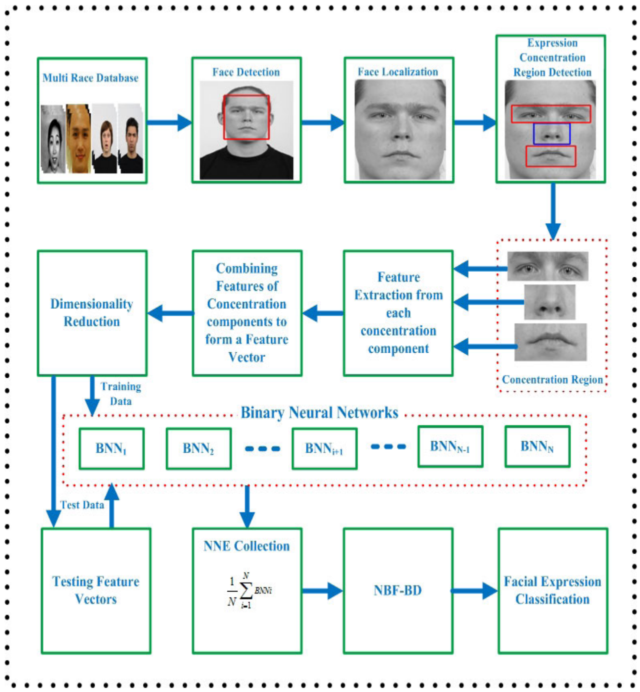


FIGURE 2. Multi-culture facial expression recognition framework Dataset Preparation.

IV. METHODOLOGY

The goal of our research is to develop an automated facial expression classification system that can classify the five universal expressions, sadness, happiness, anger, fear, and surprise. The proposed system is divided into the following four phases.

- Preparation of dataset for the training of classifiers.
- Feature extraction and dimensionality reduction from the preprocessed dataset.

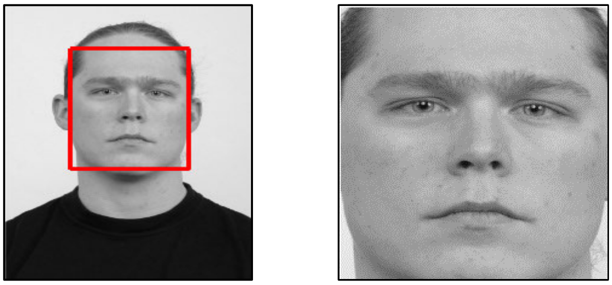


FIGURE 3. Face Detection and Localization.

- Training of classifiers using extracted feature vectors of training data.
- A results analysis technique to evaluate the performance of the proposed **approach** on the test set.

The databases used for the training and testing of classifiers are publically available and represent four different cultural, ethnic, and geographic regions. Therefore, we employed JAFFE, TFEID, and RaFD databases.

A. PREPROCESSING

Following the generation of the dataset using these databases, we applied the Viola-Jones face detector to localize the face in each image. The faces were cropped, resized, and normalized for further processing. Resizing applied using bicubic interpolation technique [33] by preserving the fine details of expressions, which produced a high-quality image in a height of 150 pixels and a width of 128 pixels. The dimensions were chosen to reduce the distortion over the whole database. Pre-processing further includes face acquisition, intensity normalization; thresh holding, and sharpening the image. Before applying the feature extraction techniques, the expression concentration region (eyes, mouth, and nose) was detected, combined, and reshaped into a single vector to represent a sample image.

An important feature of the multicultural facial expression database is the spatial uniformity of all facial images. The facial appearance varies due to many various factors such as gender, age, ethnic region, and geography [34]. Irrespective of these factors there is some common feature that defined a standardized facial structure across all cultures and ethnic regions as presented in [35]. However, these techniques focus on defining the standards for creating spatial consistency to overcome the variations of facial structure. In this research, we adopted simple techniques to preprocess the facial components. We applied Viola-Jones face detection and facial components such as eyes, nose, and mouth detection. Later, these components are combined to form the region of interest which is used for feature extraction.

B. FEATURE EXTRACTION

Feature extraction is of fundamental importance in the classification of facial expression process. It plays a pivotal role in increasing the classifiers' accuracy. It is significant to extract the features which are robust to illumination, noise, and face alignment. Subsequently, the process of dimensionality reduction is also very important because the feature extraction process extracts many irrelevant features, which may affect the accuracy of the classifier. However, it is necessary to obtain the most significant features which contribute most to the expression representation. Feature extraction and dimensionality reduction process was carried out using PCA, ULBP, and LBP. As various feature extraction techniques extract numerous features that may result in increased computational cost and complexity. Hence, extracted features are reduced using principal component analysis [138], which is a widely used dimension reduction technique.

1) LOCAL BINARY PATTERN

The local binary pattern (LBP) descriptor was initially reported by Ojala *et al.* [36]. Since then, it has drawn the attention of the researchers working in facial expression and face recognition. It has proved to be a very important facial expression recognition technique by comparing the pixel intensity of neighboring pixels with the center pixels of the 3×3 window.

$$LBP(x_c, y_c) = \sum_{n=0}^{7} (P_n - P_c)^{2^n} \quad (1)$$

2) UNIFORM LOCAL BINARY PATTERNS

Ojala *et al.* in [36] presented an extension of the LBP operator with different sizes of the neighborhood to capture the discriminate features at different scales. In this technique, a circular local neighborhood is defined as a set of neighboring points using bilinear interpolation. It allows defining a neighborhood of any radius with any number of neighboring pixels. It defines the texture of the image at different scales which helps to extract the micro patterns more efficiently.

$$LBP_{P,R} = \sum_{i=0}^{P-1} S(x_i - x_c)2^{\mathrm{j}} \quad (2)$$

Here x_i is the neighbor pixel of central pixel $\mathrm{x_c}$. The notation (P, R) represents a circular neighborhood of P pixels on a circle of radius R. The number of neighborhoods P and radius R can be extended to compute large scale texture descriptors according to requirements. The operator LBP P, R produces 2P different number of histogram bins with different values of P.

For 8 points, the numbers of histogram bins are 256, and 4096 for 12 points. The circular LBP operator is not invariant to image rotation because after rotation the neighboring pixels will move accordingly along the perimeter of the circle, which will produce a different LBP code. To handle these rotational effects a rotation-invariant LBP descriptor is introduced in [108]. The rotation invariant LBP operator works as follows:

$$LBP_{P,R} = \min \left\{ROR(LBP_{P,R,i})\right\} \quad i = 0, 1, \ldots, P-1 \quad (3)$$

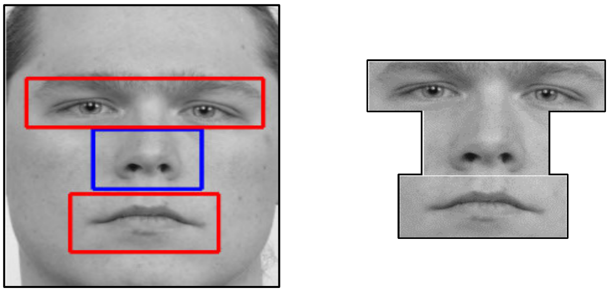


FIGURE 4. Expression concentration Region detection.

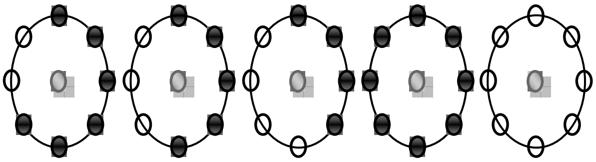


FIGURE 5. Circular pattern with different micropatterns (left to the right line, corner, edge, spot, and flat).

Here ROR (x, i) rotates each bit of binary string of length P, x, i times. The LBP (P, R) operator counts the occurrence of each rotational pattern which belongs to a microfeature in the facial image. Therefore, these feature vectors can be used to extract the micro details from facial images. The circular pattern detects many different micro patterns such as a corner, edge, line, flat, and spot. These patterns are combined to represent local texture information. Fig. 4 shows the results of detected circular patterns like line, corner, edge, spot, and flat.

The subset of patterns is called uniform local binary patterns [108]. A uniform LBP pattern contains a maximum of two bitwise shifts from 1 to 0 and 0 to 1, like 11011111, 11110001, and 00110000. Where 01001101, 11101000, and 00001011 are examples of the non-uniform pattern.

$$LBP_{P,R} = \sum_{i=0}^{P-1} S(x_i - x_c) 2_u^{j2} \tag{4}$$

$$S(x) = \begin{cases} 1 & if\,(x \geq 0) \\ 0 & if\,(x < 0) \end{cases} \tag{5}$$

LBPu2 (P, R) denotes the uniform pattern, where u2 is used to label the uniform patterns and non-uniform patterns are labeled with a single label. The total number of binary patterns in uniform LBP is computed as (P − 1) P + 2. Where (P − 1) P are the rotational invariant patterns, including lines, edges and corners, spot, and flat patterns. The LBP (8, 1) represents the texture features in 256 histogram bins. Whereas, LBPU2 yields a lesser number of patterns as compared to the LBP operator.

For example, uniform LBPU2 (8, 1) represents the facial image in 58 uniform feature vectors and LBPU2 (12, 1.5) represents the texture information in 134 feature vectors.

Fig. 5 illustrates the LBPU2 representation of u patterns. The points black and white represent the binary value as 1 and 0 respectively. There are eight possible representations of each uniform pattern except flat and spot. Experiments show that the texture representation of uniform LBP is similar to standard BP with a smaller number of features. Fig. 5 represents the eight rotational patterns called uniform patterns.

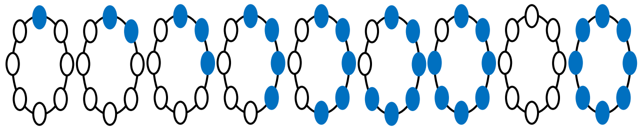


FIGURE 6. Rotational patterns of the uniform local binary pattern.

3) Principle Component Analysis (PCA)

Principle Component Analysis (PCA) is employed to explore important facial aspects for expression classification. These aspects represent the whole facial image with a subset of principal components called optimal Eigenvectors. It is an efficient dimension reduction technique to extract a minimal set of features from a large feature set without losing the important features.

Let training set X consists of N facial expression images with X = {x_1, x_2 … x_n}, then the mean image Ψ is defined as:

$$\Psi = \frac{1}{N} \sum_{i=1}^{N} X_i \tag{6}$$

Here X_i is the m x n dimension vector. The covariance matrix C is defined as:

$$C = AA^T \tag{7}$$

$A = \{\Phi_1, \Phi_2, \Phi_3, \ldots, \Phi_n\}$ is the Eigenvector matrix of C and $\Phi_i = X_i - \Psi$ is the difference of the X_i image vector from the mean image. The dimensionality can further be reduced by:

$$C = A^T A \tag{8}$$

C. ARTIFICIAL NEURAL NETWORK ENSEMBLE

The process of constructing an ANNE classifier is categorized into three phases. The first phase is to train a set of base-level classifiers for each expression against all other expression samples. Whereas the 2^{nd} phase is to construct the ANN collections by combining the results of base-level binary classifiers using (12). At last, the decisions of n ANNE collections are combined with a final predictor to recognize the presence of an expression.

1) BINARY NEURAL NETWORK

It is a feed-forward network with a tan-sigmoid transfer function that can be employed to excite the neurons to generate the output value of each neuron in hidden as well as the output layer. The output layer has one neuron with two classes associated with each input vector. The output neuron represents a single class either 1 or 0. The binary neural network

Algorithm 1

Begin: Training samples $\{d_1, e_1\}, \ldots, \{D_n, E_n\}$;
d_i is the sample dataset and e_i is the expression;
Repeat e = 1 to E **do**
Randomly select subset S_E and V_E from training set D
Repeat n = 1 to N **do** (Number of ANNEs)
Repeat k = 1 to K **do** (Number of BNN in each ANNE)
Randomly select a subset d_i from S_E and v_i from V_E
BNN (d_i, v_i, e_i)
End
Combine the hypothesis h_i of BNNs in an ANNE

$$P_{(x)} = \frac{1}{K}\sum_{i=1}^{K} hi$$

End
End
Combine the decisions of ANNEs
Naive Bayes Predictor
Construct meta-data using training set D

$$Y(d) = \max p(y) \prod_{d} p_{y_x}^{d_x} (1 - p_{yx})^{1 - d_x}$$

End

can classify any unknown sample for prevalence within a particular expression. The output of the binary neural network represents the presence of one of the facial expressions.

2) ARTIFICIAL NEURAL NETWORK ENSEMBLES STRUCTURE

To combine the decisions of ANNE collections we adopted the stacking ensemble technique. However, NBBD was introduced as a final predictor in an adopted stacking ensemble approach. Hence, the NB classifier as the final predictor learns from ensemble collection's predictions to minimize the error of base-level classifiers.

The major objective of this research is to learn the errors of base-level classifiers to recognize the multicultural facial expression efficiently. During the 1st, a set C = {C1, C2...Ck} of K represents the number of base-level binary classifiers (BNN1, BNN2...BNN_k). A random vector Vi is generated (for i-th BNN) independent of the previously reported random vectors V1...Vi-1 but with the same distribution. A new BNN is trained using the expression specific training set S randomly selected from SE and Vi, resulting in a BNN (d_i, v_i, e_i). Here d_i represents expression specific input vector, e_i represents class whereas v_i represents other expression input vector. In a random selection of input vector V, it is comprised of a no. of independent vectors with random features (from 1 to N). The number of samples and dimensionality of V depends on its use in the BNN training protocol (see Table 3).

After training of N number of BNNs, their outputs are combined by using (9) and (10), to determine the prevalence of a pose dependent expression. The resultant classifier is called ANNE.

To recognize an unknown facial expression x the N decisions $\{ANNE_{1(x)}, ANNE_{2(x)}, \ldots ANNE_{n(x)}\}$ of the ANNE collections in each ensemble collection (Multiple ANNEs for each expression) on x are trained. The decisions of these ANNE collections are then used to form the metadata to train the final predictor. A pool of ANNEs trained for each expression to recognize the cross-culture expression. The feature vector of the training set is further evaluated by ANNEs to produce a meta-vector. This meta-vector is the input to an NBBD classifier that predicts the prevalence of one of the five expressions. ANNEs' decision depends on the output of base-level classifiers (BNNs) to compute the probability to determine the prevalence of an expression. The resulting ANNE collections give accuracy that compares favorably with state-of-the-art ensemble techniques.

$$P_{(x)} = \frac{1}{K}\sum_{i=1}^{K} hi \tag{9}$$

$$S(x) = \begin{cases} 1 & if\ (P(x) \geq \theta) \\ 0 & if\ (P(x) < \theta) \end{cases} \tag{10}$$

3) ARTIFICIAL NEURAL NETWORK ENSEMBLE TRAINING

The extracted features with reduced dimensionality are represented in a feature vector to train to the binary classifiers. Then the decision of binary classifiers is combined the form an ANNE collection. The decisions of ANNE collections generate the meta feature vector, which is used to train the NBBD predictor. During testing, this meta feature vector is the input to an NBBD predictor, that predicts the presence or absence of expression. ANNEs, being classifiers relies on the output of BNNs to decide eh presence of an expression.

The procedure of ANNE collections constructions is presented in Algorithm 1, which starts from a random selection of feature vectors from training feature vectors, followed by a random selection of subset feature vector, from expression specific training data, for a particular ANNE. The number of ANNEs was varied to be 50 (10 for each expression), 100 (20 for each expression), 150 (30 for each expression), 200 (40 for each expression), 250 (50), or 300 (60). The final decision about the number of ANNEs was decided based on the average accuracy of the ensemble classifier on the test set. Therefore, each expression recognized using 60 ANNEs, in the case of 300 ANNEs across the classifier.

Three different types of ANNE collections were trained using Table 3 parameters with different training arrangements.

- In each ANNE collection, the BNNs were trained using subset (culture-specific) feature vectors of each expression against all other cultural feature vectors of the whole dataset. This technique tries to learn the cultural variation in representing the facial expressions. It does not try to emphasize intra cultural expression variability.

- The BNNs trained using feature vectors of each expression (including feature vectors of all cultural expressions) against all other expressions feature vectors. This model learns cross-cultural expression similarity for the same expression. It also learns inter-expression variability in the case of different expressions as described in [37], [38].
- The classifiers trained using only culturally specific feature vectors to learn intra-cultural expression variability and intracultural similarity in each expression.

The predictions of each ANNE collection were computed with the probability of the decisions of base-level classifiers.

$$P(x) = \frac{1}{N}\sum_{i=1}^{N} hi \quad (11)$$

$$S(x) = \begin{cases} 1 & if\ (x \geq \theta) \\ 0 & if\ (x < \theta) \end{cases} \quad (12)$$

4) NAIVE BAYS WITH BERNOULLI DISTRIBUTION

Naïve Bays with Bernoulli distribution (NBBD) was introduced as a final predictor to learn about the presence of an expression from the output of ANNEs. The NB classifier is a conditional probability-based classification technique. Naive Bayes performed remarkably as compare to many very well in practice, often comparable to many modern classifiers.

$$c(d) = \max \frac{1}{Z(d)} p(c) \prod_{i=1}^{I} p(d_i|c) \quad (13)$$

Here Z (d) is a scale parameter that depends on d (D-dimensional binary vector). ANNEs are employed as a binary classifier that predicts about the presence or absence of expression. While the binary decisions in this research we introduced the fusion of NB with Bernoulli distribution (BD). We believe that the fusion of NB with BD operates significantly as compared to traditional NB classifier.

$$c(d) = \max p(c) \prod_{d} p_{cx}^{dx} (1 - p_{cx})^{1-dx} \quad (14)$$

Prior probabilities were computed while considering the predictions of ANNEs on training data. Hence ANNEs errors are minimized by considering the performances of BNNs (on training data).

D. MODEL PERFORMANCE EVALUATION

The most appropriate combination of classifiers (ANNE, BNN, and NB) training parameters are selected based on performance on the training set. The method used was the frequency of accuracy over training data. For each of the three ensemble classifier schemes, 20 BNNs were trained for three ANNE collection types. The ensemble classifier with the highest accuracy was used to choose the best expression recognition combination of classifiers. Experimental results depicted that performance of ensemble classifiers with 200, 250 and 300 ANNE collections was significant as compared with those of 50, 100, and 150 ANNEs. Anyhow 300 ANNEs appeared to be the best using PCA feature vectors. ANNE collection parameters determined for LBP, PCA, and ULBP features are given in Table 1.

TABLE 1. Optimal Classifier Structure for PCA, LBP, and ULP Features.

Parameters	PCA	LBP	ULBP
ANNE collection type	2	2	2
No of BNNs per ANNE	20	20	20
No ANNE collections	300	150	250
Accuracy	90.1	79.5	80.0

V. EXPERIMENTS

The experiments performed on the multi-culture database (produced by combining the RadBoud, TFEID, and JAFFE databases) to evaluate the performance of the proposed multi-culture facial expression classification framework. This database holds a total of 1386 images, where 49 female participants and 58 male participants. Moreover, it contains 838 images from the RadBoud faces database (19 female and 38 male participants), 398 images from TFEID (20 female and 20 male participants), it also contains 150 images from JAFFE (10 female actresses). The classification process was carried out using all three types of ANNE collections by varying the number of ANNEs (100, 150, 200, 250, and 300) with 20 BNNs. The NBBD ensemble classifier was used to determine the presence of expression from ANNE collections' outputs. The complete evaluation results are represented in Table 2, Table 3, and Table 4, along with accuracy on each expression.

Table 2, Table 3, and Table 4 represent the complete experimental results of ANNE collections type 2, type 1, and type 3. Each ANNE collection consists of 20 BNN classifiers. In type 1 ANNE collections, the BNNs were trained using subset (culture dependent) feature vectors of each expression against all other cultural feature vectors of the whole dataset. Moreover, in type 2 BNNs were trained using feature vectors of each expression (including feature vectors of all cultural) against all other expressions feature vectors. The type 3 classifiers trained using only culturally specific feature vectors.

The whole test data is used for the evaluation of three types of ANNE collections. From experimental results, it can be observed that the ANNE collection combines with NBBD using the PCA feature vector outperforms the ANNE collections LBP and ULBP feature. The best average facial expression recognition accuracy for the NBBD ensemble classifier is 90.14% using PCA feature vectors.

From Tables 2, 3, 4 we can observe that performance of the NBBD ensemble classifier is superior on Eigen features as compare to LBP and LBP. However, expression recognition accuracy for expression fear and sadness is very low using

TABLE 2. Neural Network Ensemble Collection Type 1 Accuracy (%) with NB, KNN, and SVM Ensemble Classifiers.

ANNEs Count	300	250	200	150	100
PCA					
Anger	53.65	55.26	50.00	48.80	58.42
Happiness	64.36	58.62	**94.66**	61.53	50.68
Surprise	62.82	58.62	59.21	60.00	63.85
Sadness	53.40	59.25	54.11	56.96	60.71
Fear	71.60	77.64	45.19	62.19	60.91
Average	61.06	**62.02**	59.38	57.93	59.13
LBP					
Anger	60.60	67.90	66.19	73.49	58.53
Happiness	70.32	72.72	70.75	70.73	72.00
Surprise	**71.42**	67.74	64.38	64.44	71.26
Sadness	47.61	49.45	**52.22**	55.68	43.82
Fear	**85.71**	72.34	63.15	60.27	61.44
Average	**67.79**	65.87	63.46	64.90	61.06
ULBP					
Anger	61.61	64.04	63.63	69.33	60.75
Happiness	76.82	73.17	69.23	72.09	73.61
Surprise	**66.66**	64.21	63.76	70.12	63.09
Sadness	49.35	64.10	65.16	51.25	60.21
Fear	61.25	77.77	**71.11**	64.28	64.77
Average	63.22	**68.27**	66.83	65.38	64.18

TABLE 3. Neural Network Ensemble Collection Type 2 Accuracy (%) with NB, KNN, and SVM Ensemble Classifiers.

ANNEs Count	300	250	200	150	100
PCA					
Anger	90.24	86.58	86.58	87.80	90.24
Happiness	95.40	97.70	97.70	97.70	97.70
Surprise	98.70	93.50	93.50	93.50	96.10
Sadness	84.94	76.34	77.41	76.34	74.19
Fear	81.81	84.41	81.81	81.81	81.81
Average	**90.1**	87.50	87.26	87.26	87.74
LBP					
Anger	92.68	89.02	89.02	91.46	90.24
Happiness	89.65	88.50	88.50	93.10	89.65
Surprise	88.31	85.71	85.71	88.31	85.71
Sadness	61.29	44.08	46.23	62.36	43.01
Fear	66.23	53.24	53.24	63.63	53.24
Average	79.33	71.63	72.12	**79.57**	71.88
ULBP					
Anger	92.68	92.68	92.68	91.46	92.68
Happiness	89.65	89.65	88.50	88.50	93.10
Surprise	88.31	88.31	88.31	87.01	89.61
Sadness	61.29	62.36	62.36	61.29	61.29
Fear	64.93	68.83	64.93	68.83	62.33
Average	79.09	**80.05**	79.09	79.09	79.57

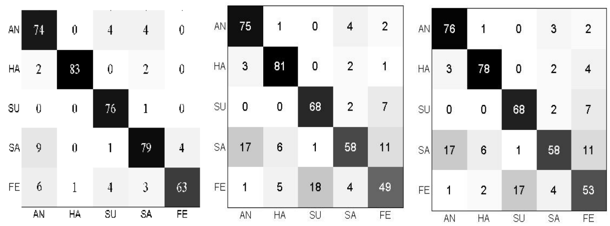
LBP and ULP features. Therefore, accuracy difference is very high for PCA and other features on expressions fear and sadness, the difference is about 32% on sadness and about 25% on fear.

TABLE 4. Neural Network Ensemble Collection Type 3 Accuracy (%) with NB, KNN, and SVM Ensemble Classifiers.

ANNEs Count	300	250	200	150	100
PCA					
Anger	75.60	70.52	74.48	78.20	70.37
Happiness	93.75	93.50	90.24	87.05	95.00
Surprise	91.20	89.13	93.15	91.01	91.11
Sadness	80.00	86.66	85.33	64.63	86.02
Fear	64.77	77.92	80.68	80.48	87.50
Average	81.01	83.17	84.13	80.53	**86.06**
LBP					
Anger	73.75	81.48	77.63	76.66	77.92
Happiness	89.87	79.51	86.36	85.52	88.75
Surprise	92.40	83.95	89.15	95.23	83.75
Sadness	68.13	71.26	72.94	66.66	64.89
Fear	66.66	66.66	72.61	69.62	70.58
Average	77.64	76.44	**79.81**	78.61	76.68
ULBP					
Anger	76.54	76.34	76.82	73.75	74.41
Happiness	87.32	87.34	85.85	92.95	85.41
Surprise	82.35	76.25	81.33	86.58	87.65
Sadness	71.91	73.86	64.19	70.21	74.66
Fear	73.33	68.42	68.35	68.53	80.76
Average	77.88	76.44	75.72	77.64	**80.77**

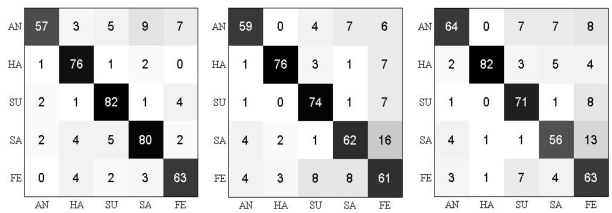
The confusion matrices presented in Fig. 6, Fig. 7, and Fig. 8 demonstrates the best expression classification accuracies represented in Tables 2, 3, and 4. These confusion matrices representing the performance of three types of ensemble classifiers with three types of feature vectors (PCA, LBP, and ULBP).

The greyscale color-map matrix is designed to make simpler the representation of experimental results, where correctly classified samples are presented as black and the intensity of misclassified samples is presented with grayscale. The off-diagonal values are representing the misclassified samples and diagonal values are representing correctly classified samples. The intensity of off-diagonal values highlights the generalization of the proposed ANNE collections. The confusion matrices are generated by putting the actual facial expression label on the x-axis and ANNEs predicted label on the y-axis. Each row is representing the level of difficulty in recognizing a facial expression. The intensity of off-diagonal values representing the level of resemblance between two expressions, higher intensity means higher resemblance, lower intensity representing the low resemblance. Fig. 7 demonstrates the level of cross-culture expression resemblance with the highest expression recognition accuracy of 90.14%. In contrast, the expression recognition accuracies using LBP and ULBP features are very low as compare to PCA feature vectors. Only from the facial expression classification point of view, we can observe that the ensemble classifiers with Eigenvectors performing better than LBP and ULBP features. Although the performance of ensemble classifiers is remarkable on PCA, in the case of



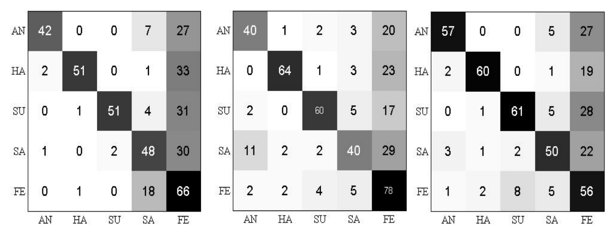
(a) PCA (b) LBP (c) ULBP

FIGURE 7. Confusion matrices of best results obtained from ANNE collections type 2 with PCA LBP and ULBP features.



(a) PCA (b) LBP (c) ULBP

FIGURE 8. Confusion matrices of best results obtained from ANNE collections type 1 with PCA LBP and ULBP features.



(a) PCA (b) LBP (c) ULBP

FIGURE 9. Confusion matrices of best results obtained from ANNE collections type 3 with PCA LBP and ULBP features.

ANNE collection type 1 the accuracy on LBP and ULBP features is higher than PCA features.

As Table 3 shows, the best expression recognition accuracy achieved with 300 ANNEs, where the Eigen feature vectors used to train the ANNE collection. It demonstrates that 300 ANNEs dominating the 250, 200, 150, and 100 ANNEs. On the other hand, while ANNEs trained using LBP and ULBP feature vectors the expression recognition accuracy with 300 ANNEs is very low ac compare to 50, 100, and 150 ANNEs. From Tables 2, 3, and 4, we can observe the highest expression classification accuracy achieved using LBP and ULBP descriptors with 150 and 100 ANNEs respectively.

Fig. 8 demonstrates that anger and fear expression have a high misclassification rate as compared to other expressions. The reason behind this is that binary neural networks were trained with culture-specific facial images and tested with the multicultural dataset. The binary neural networks were unable to learn the cross-culture facial variations. It misclassifies those expressions which have a high resemblance rate due to cultural variations because the muscular deformation of both facial expressions is similar in different cultures. We can

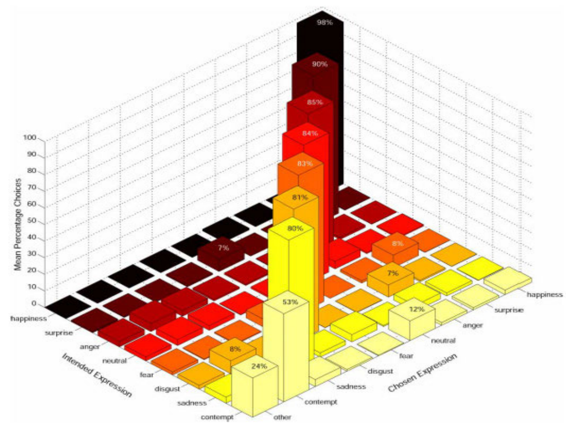


FIGURE 10. RadBoud human recognition accuracy [39].

also see that the misclassification rate is very high in case of happy and fear expressions. These results also demonstrate the fear is the most confusing expression and surprise is the least confusing expression as compared to other expressions. Next to fear and anger, happiness and sadness are the most confusing facial expressions.

These confusion matrices show that fear and sadness expression have a high misclassification rate as compared to other expression combinations. On the other hand, the facial expressions of surprise and happiness have a high true classification rate. While observing the highest expression recognition accuracy from all ANNE types, happiness has the highest recognition rate and lowest misclassification rate. From the best performing ANNE, the expression of surprise has the highest expression recognition accuracy of 98.70%.

VI. ANALYSIS

Here we focus on the best performances achieved from experimental results. The results demonstrate that the performance of NNEs is not the same on all expressions. When we analyze the results presented in Fig. 10 we can see that the human recognition accuracies for RadBoud [39] face database, the highest recognition accuracy of expression surprise, and happiness is 90% and 98% respectively. These two expressions are easy to recognize as compare to other expressions, which is also proved by results achieved with ANNEs. It also indicates that the expression of fear and sadness is harder to distinguish accurately, which strengthens the correctness the results achieved using the proposed approach.

We also evaluated the performance of ANNE collections by using facial images of different cultures for the training and testing of binary neural networks. The experimental results on facial images of different cultures for training and testing are presented in Table 5. These results highlight that the proposed framework performed significantly even when using facial images of different cultures. However, its performance is slightly lower than the best expression recognition presented in Table 3. The best expression recognition is 90.1% while using the multicultural dataset. Whereas,

TABLE 5. Best Ensemble Classifiers Trained and Tested with Facial Images of Different Cultures.

Test Database	Training Database	ANNEs	Accuracy
JAFFE	TFEID, RadBoud	100	86.67
TFEID	JAFFE, RadBoud	150	83.19
MOROCCAN	JAFFE, TFEID, Caucasian	150	89.47
CAUCASIAN	JAFFE, TFEID, Moroccan	300	86.36

the facial expression recognition accuracy is 89.47% when the classifiers tested on facial images of Moroccan culture, and trained using the facial images of Japanese, Taiwanese and Caucasian cultures. It demonstrates that the proposed framework also performed significantly when the classifiers trained and tested using the facial images of different cultures. Again, this consistency in expression recognition accuracy provides evidence about the generalization of ANNE collections. An interesting additional aspect could be that humans can efficiently recognize the facial expressions of unknown subjects even in crowdsensing [43], [44].

It indicates that the variations in facial appearance do not influence on humans' ability to classify facial expressions accurately. The results present in Table 5 also supports these arguments that the proposed technique performing better on unseen samples. In [40] it is presented that facial expression recognition accuracy of 50% is significant when classifiers are trained and tested on facial images of different cultures. However, the classification accuracy presented in Table 5 is higher than 80%, which proves the correctness of ANNE collections. These results show that ANNE collections perform consistently even when facial images of different cultures are used for the training and testing of classifiers.

According to the literature presented in Table 6, where facial images of one culture are used for the training of the classifier, and the facial images of other cultures are used for the testing of the classifier. During the training of BNN, we observed that different combinations of parameters are required when using the different datasets for the training of the same classifier. Because the facial databases are naturally diverse, this is impossible for a classifier to learn the inherited variations in different databases. Therefore, it is required to develop such techniques which can learn the cultural variation to recognize the multicultural facial expressions. In this research, the effect of cultural variations is minimized while focusing on the expression concentration region. Moreover, the introduction of multiple BNNs at the base level also minimizes the effect of variation in facial expression representation and facial structures.

A. PERFORMANCE EVALUATION WITH CROSS-VALIDATION

To generalize the performance of the proposed ensemble technique we have applied 10-fold cross-validation. With 10-fold cross-validation, we achieved average expression recognition accuracy of 89.31% with 300 ANNEs trained using PCA feature vectors and NB predictor. These results demonstrate the generalization of the proposed technique by varying the training and test set.

TABLE 6. Existing Methods' Accuracy when Facial Images of Different Cultures are used for the Training and Testing of Binary Neural Networks.

Method	Training Database	Test Database	Accuracy
2015 [8]	CK+	JAFFE	42.3
2015 [8]	CK+	MUG	47.8
2015 [8]	CK+	BOSPHORUS	43.0
2015 [8]	JAFFE	CK+	48.2
2015 [8]	JAFFE	MUG	32.9
2015 [8]	JAFFE	BOSPHORUS	30.0
2015 [8]	MUG	CK+	45.6
2012 [41]	JAFFE	CK	54.05
2012 [41]	CK	JAFFE	55.87
2014 [42]	CK	JAFFE	45.19
2014 [42]	CK	TFEID	51.13
2014 [26]	BU-3DFE	KDEF	46.17
2014 [26]	BU-3DFE	RadBoud	55.03
2014 [26]	BU-3DFE	JAFFE	41.96
2014 [26]	BU-3DFE	PICS	35.81
2014 [26]	BU-3DFE	SFEW	20.57
Proposed	TFEID, RadBoud	JAFFE	86.67
Proposed	JAFFE, RadBoud	TFEID	83.19
Proposed	JAFFE, TFEID, Caucasian	MOROCCAN	89.47
Proposed	JAFFE, TFEID, Moroccan	CAUCASIAN	86.36

B. INTER-EXPRESSION RESEMBLANCE ANALYSIS

Inter-expression resemblance analysis has been performed to analyze the resemblance between different expressions, which decreases the models' performance. The expression recognition accuracy is high on those expressions which have low resemblance with other expressions. However, many studies adopted Ekman's theories to present the similarities between different expression combinations.

$$R(i,j) = \frac{C(i,j)}{\sqrt{C(i,i)C(i,j)}} \quad (15)$$

We applied the inter-expression correlation to compute the resemblance between different expression combinations. The confusion matrices presented in Fig. 6 are used to compute the correlation coefficients between the five expressions. Fig. 11 illustrates the correlation coefficients between recognized expressions. These correlations were determined using the average expression recognition accuracies of best ensemble models presented in Table 1, using (15).

From Fig. 11 we observed that there is no positive correlation between two expressions. The pair of facial expressions of sadness-anger and fear surprise has small negative correlations values which indicate that the pair of facial expressions of anger-sadness and surprise-fear looks like each other. It indicates that these expression pairs have higher resemblance as compare to other combinations of facial expressions. It demonstrates that these pairs of facial expressions have a similar facial structure. On the other hand, while considering the other combinations of facial expression a different relationship could be implied.

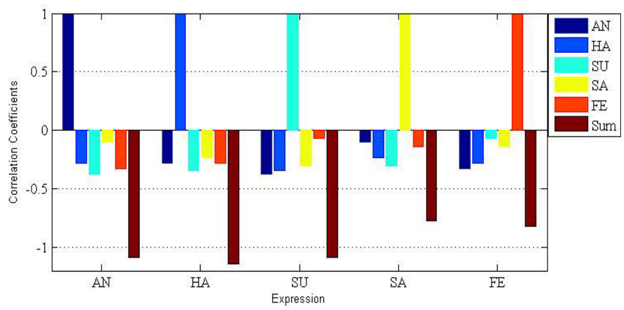


FIGURE 11. Correlation coefficients based on comparing recognized and actual expressions for each of the five expressions.

Similarly, the combination of sadness-surprise has more resemblance as compare to other expression pairs. Thus, the correlation coefficients quantify the similarity and dissimilarity between two expressions. While considering anger-surprise, which has the lowest resemblance. We can say that these two expressions have a similar facial structure and its difficult for the classifier to differentiate between these two expressions. Similarly, the expression pair happiness-surprise, anger-surprise, sadness-surprise, and fear-anger have lower resemblance as compare to other facial expression combinations.

VII. CONCLUSION

Cross-cultural expression recognition system addresses the three main problems: 1) multicultural facial expression representation variation, 2) multicultural facial structure variation, and 3) inter-expression resemblance. Due to these problems, the cross-cultural facial expression recognition is very inconsistent. To overcome the issue of multicultural variation and to represent cultural diversity, a multi-culture facial expression database is developed. The expression concentration region is detected to enable the classifiers to classify the multicultural facial expressions more accurately which also enables the classifier to learn the variations in expression representation by focusing on the expression concentration components such as eyes, nose, and mouth. To learn the cross-cultural expression representation variations, a pool of binary classifiers are trained using a multi-culture database. The binary classifiers learn cultural variations for the same expression. It also learns the inter-expression variability in case of different expressions. We observed that the use of Bernoulli distribution increased the capability of the classifier to efficiently predict the presence of an expression. There is tremendous scope of research in cross-cultural emotional recognition. The significant focus of future research will be on the development of ensemble techniques, facial feature representation techniques, and preprocessing techniques to overcome the limitations of this research.

ACKNOWLEDGMENT

(Ghulam Ali and Farman Ali are co-first authors.)

REFERENCES

[1] P. Ekman, E. R. Sorenson, and W. V. Friesen, "Pan-cultural elements in facial displays of emotion," *Science*, vol. 164, no. 3875, pp. 86–88, Apr. 1969.

[2] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion," *J. Personality Social Psychol.*, vol. 17, no. 2, p. 124, 1971.

[3] P. Ekman, "Universals and cultural differences in facial expressions of emotion," *Nebraska Symp. Motivat.*, vol. 19, no. 1, pp. 207–283, 1971.

[4] Z. Wang and N. Xiao, "Using MD-Adaboost to enhance classifier of facial expression recognition," *J. Comput. Inf. Syst.*, vol. 9, no. 3, pp. 923–932, 2013.

[5] E. Owusu, Y. Zhan, and Q. R. Mao, "A neural-AdaBoost based facial expression recognition system," *Expert Syst. Appl.*, vol. 41, no. 7, pp. 3383–3390, Jun. 2014.

[6] D. Ghimire and J. Lee, "Extreme learning machine ensemble using bagging for facial expression recognition," *J. Inf. Process. Syst.*, vol. 10, no. 3, pp. 443–458, Sep. 2014.

[7] G. Levi and T. Hassner, "Emotion recognition in the wild via convolutional neural networks and mapped binary patterns," in *Proc. ACM Int. Conf. Multimodal Interact. ICMI*, 2015, pp. 503–510.

[8] F. A. M. Da Silva and H. Pedrini, "Effects of cultural characteristics on building an emotion classifier through facial expression analysis," *J. Electron. Imag.*, vol. 24, no. 2, 2015, Art. no. 023015.

[9] M. S. Zia and M. A. Jaffar, "Cross-cultural emotion classification based on incremental learning and LBP-features," *Appl. Math. Inf. Sci.*, vol. 9, no. 5, pp. 2651–2656, 2015.

[10] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 1, pp. 39–58, Jan. 2009.

[11] S. M. Lajevardi and Z. M. Hussain, "Feature extraction for facial expression recognition based on hybrid face regions," *Adv. Electr. Comput. Eng.*, vol. 9, no. 3, pp. 63–67, 2009.

[12] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski, "Classifying facial actions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 10, pp. 974–989, Oct. 1999.

[13] X. Zhao and S. Zhang, "Facial expression recognition using local binary patterns and discriminant kernel locally linear embedding," *EURASIP J. Adv. Signal Process.*, vol. 2012, no. 1, pp. 9573–9588, Dec. 2012.

[14] T. Moeslund, A. Hilton, V. Kruger, and L. Sigal, Eds., *Visual Analysis of Humans: Looking at People*. Springer-Verlag, 2011.

[15] B. Jayanta, B. Debnath, and T. Kim, "Use of artificial neural network in pattern recognition," *Int. J. Softw. Eng. Appl.*, vol. 4, no. 2, pp. 23–34, 2010.

[16] L. C. Jain and L. I. Kuncheva, "Designing classifier fusion systems by genetic algorithms," *IEEE Trans. Evol. Comput.*, vol. 4, no. 4, pp. 327–336, Nov. 2000.

[17] B. Minaei-Bidgoli, G. Kortemeyer, and W. Punch, "Optimizing classification ensembles via a genetic algorithm for a web-based educational system," in *Proc. Int. Workshop Syntactical Struct. Pattern Recognit. Stat. Pattern Recognit.*, Berlin, Germany, 2004, pp. 397–406.

[18] H. Altinçay, "Optimal resampling and classifier prototype selection in classifier ensembles using genetic algorithms," *Pattern Anal. Appl.*, vol. 7, no. 3, pp. 285–295, Sep. 2004.

[19] F. Chen, Z. Wang, Z. Xu, J. Xiao, and G. Wang, "Facial expression recognition using wavelet transform and neural network ensemble," *Intell. Syst.*, vol. 2, no. 1, pp. 122–126, 2009.

[20] T. Windeatt, "Weighted decoding ECOC for facial action unit classification, applications of supervised and unsupervised ensemble methods," *Studies Comput. Intell.*, vol. 245, no. 1, pp. 59–77, 2009.

[21] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, nos. 1–3, pp. 489–501, Dec. 2006.

[22] Z. Y. Wang and N. F. Xiao, "Fuzzy integral-based neural network ensemble for facial expression recognition," in *Proc. Int. Conf. Comput. Inf. Syst. Ind. Appl.*, 2015, pp. 709–712.

[23] B. K. Kim, J. Roh, S. Y. Dong, and S. Y. Lee, "Hierarchical committee of deep convolutional neural networks for robust facial expression recognition," *J. Multimodal User Interface*, vol. 2016, no. 1, pp. 1–17, 2016.

[24] Y. Chen and M. Leung Wong, "An ant colony optimization approach for stacking ensemble," in *Proc. 2nd World Congr. Nature Biologically Inspired Comput. (NaBIC)*, Dec. 2010, pp. 146–151.

[25] D. Ghimire and J. Lee, "Geometric feature-based facial expression recognition in image sequences using multi-class AdaBoost and support vector machines," *Sensors*, vol. 13, no. 6, pp. 7714–7734, 2013.

[26] M. G. Villanueva and S. R. Zavala, "Deep neural network architecture: Application for facial expression recognition," *IEEE Latin Amer. Trans.*, vol. 18, no. 7, pp. 1311–1319, Jul. 2020.
[27] N. Yu and D. Bai, "Facial expression recognition by jointly partial image and deep metric learning," *IEEE Access*, vol. 8, pp. 4700–4707, 2020.
[28] M. Shi, L. Xu, and X. Chen, "A novel facial expression intelligent recognition method using improved convolutional neural network," *IEEE Access*, vol. 8, pp. 57606–57614, 2020.
[29] J. A. Russell, "Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies," *Psychol. Bull.*, vol. 115, no. 1, pp. 102–141, 1994.
[30] H. A. Elfenbein, M. K. Mandal, N. Ambady, S. Harizuka, and S. Kumar, "Cross-cultural patterns in emotion recognition: Highlighting design and analytical techniques.," *Emotion*, vol. 2, no. 1, pp. 75–84, 2002.
[31] D. Matsumoto, A. Olide, J. Schug, B. Willingham, and M. Callan, "Cross-cultural judgments of spontaneous facial expressions of emotion," *J. Nonverbal Behav.*, vol. 33, no. 4, pp. 213–238, Dec. 2009.
[32] M. N. Dailey, C. Joyce, M. J. Lyons, M. Kamachi, H. Ishi, J. Gyoba, and G. W. Cottrell, "Evidence and a computational explanation of cultural differences in facial expression recognition.," *Emotion*, vol. 10, no. 6, pp. 874–893, 2010.
[33] S. Yuan, M. Abe, A. Taguchi, and M. Kawamata, "High accuracy bicubic interpolation using image local features," *IEICE Trans. Fundam. Electron., Commun. Comput. Sci.*, vol. 90, no. 8, pp. 1611–1615, Aug. 2007.
[34] R. Jafri and H. R. Arabnia, "A survey of face recognition techniques," *J. Inf. Process. Syst.*, vol. 5, no. 2, pp. 41–68, 2009.
[35] B. Klare and A. K. Jain, "On a taxonomy of facial feature," in *Proc. 4th IEEE Int. Conf. Biometrics, Theory Appl. Syst.*, Washington, DC, USA, Sep. 2010, pp. 1–8.
[36] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognit.*, vol. 29, no. 1, pp. 51–59, Jan. 1996.
[37] P. Ekman and W. Friesen, *A Technique for the Measurement of Facial Movement*. Palo Alto, CA, USA: Consulting Psychologists Press, 1978.
[38] Y.-I. Tian, T. Kanade, and J. F. Cohn, "Recognizing action units for facial expression analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 2, pp. 97–115, Feb. 2001.
[39] O. Langner, R. Dotsch, G. Bijlstra, D. H. J. Wigboldus, S. T. Hawk, and A. van Knippenberg, "Presentation and validation of the radboud faces database," *Cognition Emotion*, vol. 24, no. 8, pp. 1377–1388, Dec. 2010.
[40] S. P, D. R. C. Thool, and P. D. K., "Automatic facial feature extraction and expression recognition based on neural network," *Int. J. Adv. Comput. Sci. Appl.*, vol. 2, no. 1, pp. 1–4, 2011.
[41] W. Gu, C. Xiang, Y. V. Venkatesh, D. Huang, and H. Lin, "Facial expression recognition using radial encoding of local Gabor features and classifier synthesis," *Pattern Recognit.*, vol. 45, no. 1, pp. 80–91, 2012.
[42] N. Farajzadeh, G. Pan, and Z. Wu, "Facial expression recognition based on meta probability codes," *Pattern Anal. Appl.*, vol. 17, no. 4, pp. 763–781, Nov. 2014.
[43] A. Tariq, U. Draz, S. Yasin, J. Noureen, A. Shaf, and M. Ali, "An efficient participant's selection algorithm for crowdsensing," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 1, pp. 399–404, 2018.
[44] T. Ali, J. Noureen, U. Draz, A. Shaf, S. Yasin, and M. Ayaz, "Participants ranking algorithm for crowdsensing in mobile communication," *ICST Trans. Scalable Inf. Syst.*, vol. 5, no. 16, Apr. 2018, Art. no. 154476.
[45] A. Ali, I. Yaqoob, E. Ahmed, M. Imran, K. S. Kwak, A. Ahmad, S. A. Hussain, and Z. Ali, "Channel clustering and QoS level identification scheme for multi-channel cognitive radio networks," *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 164–171, Apr. 2018.
[46] A. Ali, M. Ahmed, M. Piran, and D. Suh, "Resource optimization scheme for multimedia-enabled wireless mesh networks," *Sensors*, vol. 14, no. 8, pp. 14500–14525, Aug. 2014.
[47] M. T. Gul, A. Ali, D. K. Singh, U. Imtinan, I. Raza, S. A. Hussain, D. Y. Suh, and J.-W. Lee, "Merge-and-forward: A cooperative multimedia transmissions protocol using RaptorQ codes," *IET Commun.*, vol. 10, no. 15, pp. 1884–1895, Oct. 2016.
[48] A. Ali, L. Feng, A. K. Bashir, S. El-Sappagh, S. H. Ahmed, M. Iqbal, and G. Raja, "Quality of service provisioning for heterogeneous services in cognitive radio-enabled Internet of Things," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 1, pp. 328–342, Jan. 2020.

GHULAM ALI received the bachelor's degree in computer science from the University of Agriculture, Faisalabad, Pakistan, and the M.S. and Ph.D. degrees in computer science from the University of Central Punjab, Lahore, Pakistan. From September 2012 to May 2013, he has served as a Lecturer at the Department of Computer Science, The University of Lahore, Lahore. He also served as a Lecturer at the Department of Computer science, Government College University, from June 2013 to August 2019. He has joined the University of Okara as an Assistant Professor, in August 2019. He is the in-charge of the Department of Software Engineering in University of Okara. He has supervised 24 M.S. research students. He is supervising 12 M.S. research students and co-supervising one Ph.D. student. His research interests include image processing, human emotions analysis, machine learning, video surveillance, medical image analysis, stochastic processes, intelligent agents, and formal methods.

AMJAD ALI received the B.S. and M.S. degrees in computer science from the COMSATS Institute of Information Technology, Pakistan, in 2006 and 2008, respectively, and the Ph.D. degree in electronics and radio engineering from Kyung Hee University, South Korea, in 2015. Since 2015, he has been serving as an Assistant Professor with the Department of Computer Science, COMSATS University Islamabad–Lahore, Pakistan. From February 2018 to March 2019, he was the Postdoctoral Scholar with the UWB Wireless Communications Research Center (formerly Key National IT Research Center), Department of Information and Communication Engineering, Inha university, South Korea, and with the Mobile Network and Communications Laboratory, School of Electrical Engineering, Korea University, Seoul, South Korea. His main research interests include cognitive radio networks, multimedia communications, cloud computing, machine learning, the Internet of Things, and vehicular networks.

FARMAN ALI received the B.S. degree in computer science from the University of Peshawar, Pakistan, in 2011, the M.S. degree in computer science from Gyeongsang National University, South Korea, in 2015, and the Ph.D. degree in information and communication engineering from Inha University, South Korea, in 2018. He worked as a Postdoctoral Fellow with the UWB Wireless Communications Research Center, Inha University, from September 2018 to August 2019. He is currently an Assistant Professor with the Department of Software, Sejong University, South Korea. He has registered over four patents and has published more than 50 research articles in peer-reviewed international journals and conferences. His current research interests include sentiment analysis / opinion mining, information extraction, information retrieval, feature fusion, artificial intelligence in text mining, ontology-based recommendation systems, healthcare monitoring systems, deep learning-based data mining, fuzzy ontology, fuzzy logic, and type-2 fuzzy logic.

UMAR DRAZ received the bachelor's degree from BZ University, in 2013, the master's degree from Government College University, Faisalabad, in 2015, and the master's degree as MSCS with COMSTS University Islamabad–Sahiwal, in 2018. He is currently pursuing the Ph.D. degree with COMSATS University Islamabad–Lahore. He had been a Lecturer of computer science with the CS Department, Government College University Faisalabad–Sahiwal, from 2017 to 2019. He is also a Lecturer with the University of Sahiwal. He has distinguished number of publications in reputed journals and conferences, including best paper award from the University of La Taroob, Australia and European Alliance International (EAI) in recognition of his quality research and productivity. He was an invited reviewer of different IEEE, SCI and EAI journals and a permanent reviewer of the *Journal of Intelligent & Fuzzy Network* and different IEEE conferences. He has been a co-research partner in several projects of wireless communication and IoT. His current research interests include wireless communication, deep learning, underwater sensor networks, block-chain, and intelligent networks.

FIAZ MAJEED received the Ph.D. degree in computer sciences from the University of Engineering and Technology, Lahore, Pakistan, in 2016. He is currently an Assistant Professor with the Faculty of Computing and Information Technology, University of Gujrat (UOG), Pakistan. His research interests include data warehousing, data streams, information retrieval, social networks, and ad hoc networks. He has published more than 15 articles in refereed journals and international conference proceedings in the above areas. He is a reviewer for several renowned international journals.

SANA YASIN received the bachelor's and master's degrees from COMSATS University Islamabad–Sahiwal, in 2016 and 2018, respectively. She is currently pursuing the Ph.D. degree with COMSATS University Islamabad–Lahore, under the NRPU project An Intelligent Brain Computer Interfacing (BCI) Based Personalized Healthcare System for Neuro-Rehabilitation. She had been with Government College University Faisalabad–Sahiwal as a Lecturer of computer science with the CS Department and a Coordinator of FYP and with the Superior College, Okara, as a Visiting. She is also a Lecturer with Superior University Gold Campus, Lahore. She has distinguished number of publications in reputed journals and conferences, including best paper the University of La Taroob, Australia. She was a direct supervised more than 24 groups of FYP for the degree of computer science and software engineering and 12+ research thesis for the degree of Master of computer science. Her research interests include machine-to-machine learning (M2M), deep learning, wireless communication, brain computer interference, and underwater wireless sensor networks (UWSN).

TARIQ ALI received the Ph.D. degree in IT from University Teknologi PETRONAS, Malaysia, in 2015, and the M.S. degree in computer science from SZABIST Islamabad, Pakistan, in 2006. During M.S., his specialization area belongs to Networks and Communication. He worked as a Lecturer with the Computer Science Department, Gordon College, Rawalpindi, Pakistan, from 2007 to 2009. He has served more than two years as an IT-Manager for the IT Department of Government, Pakistan. He is currently serving as an Assistant Professor with the Computer Science Department, COMSATS University Islamabad–Sahiwal. He is on EoL leave and also doing Post Doctorate Fellowship with Najran University, Saudi Arabia. He has published a good number of journals, conferences, and book chapters in a well-known publishing platform. During Ph.D., his specialization area belongs to Underwater Wireless Sensor Networks (UWSN). His research interests include mobile and sensor networks, routing protocols, the IoT, digital communication, underwater acoustic sensor and actor networks, and machine learning algorithms.

NOMAN HAIDER received the B.S. degree in electronics engineering from Mohammad Ali Jinnah University, Pakistan, in 2011, the M.Sc. degree in engineering and information technology from the Universiti Teknologi Petronas, Malaysia, in 2014, and the Ph.D. degree in engineering and information technology from the University of Technology Sydney, Australia, and 2019. He is currently working as a Lecturer with Victoria University, Sydney, Australia. From 2012 to 2019, he has worked on multidisciplinary research projects in collaboration with professionals from academia and industry (Intel US and Intel Europe). His research interests include resource sharing and allocation for future wireless networks, network security, artificial intelligence, and context-aware learning models.