Comprehensive Study of Features for Subject-independent Emotion Recognition

A. Ashutosh
Dhirubhai Ambani Institute of Information and Communication Technology,
Gandhinagar, India
Email: ashutosh_adhikari@daiict.ac.in

R. Savitha
Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore
Email: ramasamysa@i2r.a-star.edu.sg

S. Suresh
School of computer Science and Engineering
Nanyang Technological University
Singapore
Email: ssundaram@ntu.edu.sg

Abstract— **In this paper, we conduct a comprehensive study to identify the most discriminative features that address the interpersonal variability to perform efficient human emotion recognition task. We consider three commonly used feature extraction techniques, namely, the Local Binary Patterns (LBP), the Scale-Invariant Feature Transform (SIFT) and the curvelet transforms to extract features from the images on the JAFFE data set. A subset of these features is then selected using the Double Input Symmetrical Relevance (DISR), the Conditional Mutual Information Maximization (CMIM) and the minimum Redundancy maximum Relevance (mRMR) methods. The original feature sets and the subsets are then used to train a PBL-McRBFN classifier. We conduct a subject independent study with 10 cross validations on the JAFFE data set. The average performance of the PBL-McRBFN classifier with the different feature sets and subsets are compared. In general, feature selection methods used along with the feature extraction techniques help to perform emotion recognition more efficiently. It is also observed that the subset of features selected using the mRMR on the features extracted from the SIFT technique (SIFT+mRMR+PBL-McRBFN) is the most discriminative feature subset. A statistical paired t-test also ascertains this observation. We also compare the performance of the SIFT+mRMR+PBL-McRBFN with the other results in the literature for this problem. Performance comparison shows that the SIFT+mRMR+PBL-McRBFN outperforms other state-of-the-art methods in the literature for this problem.**

Keywords—**Emotion recognition, feature extraction, feature selection, Projection Based Learning, Meta-cognitive Radial Basis Function.**

I. INTRODUCTION

Automatic human emotion recognition is one of the most challenging problems in computer vision and video analytics. However, it has a plethora of implications in areas such as lie detection, operator fatigue detection, human-computer interaction, data-driven animation, etc. Efficient computer vision and video analytics are possible through emotion recognition through body gestures [1], speech [2] or facial expressions [3]. Earlier, it has been shown that facial expressions have a much greater impact on a listening interlocutor than voice intonation and actual spoken words [3]. Thereafter, emotion recognition from facial expression is gaining widespread interest.

The task of automatic human emotion recognition from facial expressions can be divided into two important components, namely, emotion representation and emotion classification. Emotion representation includes extracting significant features from the raw images. Feature representation involves feature extraction and feature reduction techniques. Various researchers have attempted to solve both these components, either together or separately. Some of the popular techniques include Gabor filters [10], Local Binary Patterns [5], Scale Invariant Feature Transform (SIFT) [7], Curvelet transforms [8], etc. In [4], authors have utilized Local Binary Patterns [5] for extracting features from facial images associated with Support Vector Machines (SVM) for the classification of emotions. Kullback Liebler and weighted majority classifier has been employed for classifying key point descriptors for facial images generated by a variant of Scale Invariant Feature Transform [6]. Discriminative SIFT has been deployed for finding keypoints from facial images for person-independent facial expression recognition [7]. In [8], Local Binary Patterns (LBP) and Curvelet extraction methods have been combined for efficient facial emotion representation. There are many other popular methods such as FACS (Facial Action Coding Systems) [9] which required human operators to observe face to analyze the emotions. Also, Gabor 2D Filters have been employed for extracting features from facial images [10]. K-Nearest Neighbor classifier has been employed for facial expression recognition in [11] along with Gabor filter based facial features extracted by using five frequencies and eight orientations.

As feature extraction might result in highly correlated features, it is imperative to apply feature selection methods to identify the most discriminating features. Feature selection can be done through a filter-based method or a wrapper-based method. Wrapper based methods uses the performance of a pre-determined classifier to choose features that result in superior performance of the learning model used. Thus, the wrapper based methods are computationally expensive, when the number of features is large, as in the case of Human emotion representation. Filter based methods, on the other hand, is data-dependent and select features based on the characteristics of the features in the training data. Therefore, filter based methods are desirable in such applications. Another important component of emotion recognition is emotion classification and many decision making tools like linear discriminant analysis (LDA) [17, 18], linear programming [19] and SVMs [20, 21] have been employed for learning emotion recognition tasks. However, all these classifiers learn from all the samples in the training set, as they are presented to them. Thus, they implicitly assume *a priori* uniform distribution of samples, which cannot be guaranteed. Recently, meta-cognitive learning algorithms have been adapted into machine learning literature [23-27] from human learning [22]. Meta-cognition refers to knowledge about knowledge. In machine learning framework, meta-cognition is enabled through self-regulated learning, where the learner decides *what-to-learn*, *when-to-learn* and *how-to-learn* from the training sample set. Meta-cognitive learning algorithms are shown to have better generalization ability [23-27] owing to their ability to learn evenly across the sample space. Hence, in this study, we use the Projection Based Learning algorithm of a Meta-cognitive Radial Basis Function Network (PBL-McRBFN) classifier.

Moreover, although there are several studies on the feature extraction and feature selection methods for emotion recognition, it is important to conduct a comprehensive study on the various feature extraction and feature selection methods to identify the most discriminating feature set for this task. Another major difficulty in an emotion recognition task arises due to Intra/Inter-personal variability in expressing emotions. Subject independent analysis is imperative to address the problem of intra/inter-personal variability while expressing emotions. There are a few studies in the literature that address the problem of subject independent emotion recognition [12], [13], [14], [15].

In this paper, we conduct a comprehensive study of the most commonly used feature sets in solving the human emotion recognition task, while simultaneously addressing the interpersonal variability in expressing emotions. In doing so, we use important feature extraction methods for emotion recognition, namely, the Local Binary Patterns (LBP) [5], the Scale Invariant Feature Transform (SIFT) [7], the curvelet transform [8], and study their performances in classifying emotions. As some of these features may be redundant, we use the most commonly used information theoretic filter-based feature selection techniques, namely, Conditional Maximum Information Maximization (CMIM) [44], Minimum Redundancy Maximum Relevance (mRMR) techniques [16, 38], and Double Input Symmetrical Relevance (DISR) [40], to select relevant features from the features extracted through LBP, SIFT and the curvelet transform methods. Next, we use the features extracted and selected through the aforementioned techniques to classify human emotions from the publicly available JAFFE data set. Earlier, Projection Based Learning - Meta cognitive Radial Basis Function Neural Networks (PBL-McRBFN) [27] has been applied to solve the problem of human action recognition in [28], and it performs subject independent action recognition efficiently. Therefore, in this study, we employ the PBL-McRBFN [27, 28] as the classifier to solve the subject independent human emotion recognition task. The PBL-McRBFN uses a meta-cognitive learning algorithm to decide *what-to-learn*, *when-to-learn* and *how-to-learn* using a self-regulatory learning mechanism.

The performance study shows that the features selected by the mRMR method, from the features extracted through the SIFT method are the most discriminating features for subject-independent human emotion recognition. A statistical study using paired t-test also confirms this observation. We also compare the performances of the PBL-McRBFN with the most distinguishing features with that of the best reported results in the literature for this data set. Performance studies show the superior emotion recognition ability of the classifier.

The rest of the paper is organized as follows: Section II briefly presents the description of the JAFFE data set, the various feature extraction methods, the various feature selection methods and the learning algorithm of the PBL-McRBFN classifier. The performances of the PBL-McRBFN classifier with the various feature sets are presented in Section III. Section III also presents the performance comparison of the PBL-McRBFN classifier using the most discriminating feature set with the other best performing classifiers for the problem. Finally, Section IV summarizes the findings of the study and presents the concluding remarks.

II. MATERIALS AND METHODS

In this section, we present the data set and the algorithms that are used for emotion recognition. We first briefly describe the data set used in this study. Next, we describe the various feature extraction methods, namely, Local Binary Patterns (LBP), Scale invariant Feature Transform (SIFT), and Curvelet transform methods that are used to obtain features to represent facial expression. Then, we present the feature selection methods, namely, Conditional Mutual Information Maximization (CMIM), Double Input Symmetrical Relevance(DISR), and the Minimal Redundancy Maximal Relevance (mRMR) methods that are used in this study. Finally, the Projection Based Learning algorithm of the Meta-cognitive Radial Basis Function (PBL-McRBFN) classifier is explained briefly.

Materials: JAFFE Dataset

We use the Japanese Female Facial Expressions (JAFFE) database [29] in our study to conduct a complete study on the feature extraction and feature selection methods on emotion recognition problem. The data set consists of 213 images posed by 10 female models, categorized into 7 different emotions. Thus, the problem of emotion recognition is defined as a classification problem of classifying the facial expressions into one of 7 classes. The problem can be solved with sufficiently discriminating features and classifier. The objective of this study is to identify these discriminating features through a study across several feature extraction and feature selection methods to solve this problem.

The various feature extraction and feature selection methods are discussed briefly in this section. Fig. 1 shows an overview of the process of emotion recognition from images of facial expressions. From the figure, it can be seen that the process involves feature extraction, feature selection and a classifier. In this work, we use the SIFT, LBP and curvelet transforms as the feature extraction methods, the mRMR, DISR and CMIM as the feature selection methods and a Projection Based Learning algorithm of a Meta-cognitive Radial Basis Function classifier (PBL-McRBFN) as the classifier. We briefly describe all these methods briefly in this section.

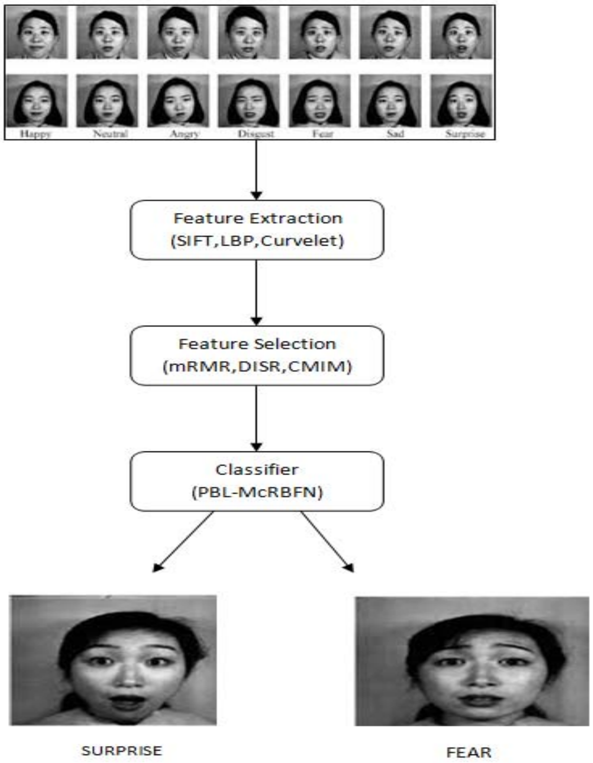

Figure 1: *Overview of the Process of Emotion Recognition from Facial Expressions*

A. Feature Extraction Methods

In this section, we briefly describe the various feature extraction methods we use for the emotion recognition problem, namely, the Local Binary Patterns (LBP), Scale Invariant Feature Transform (SIFT), and Curvelet Transform.

a. **Local Binary patterns (LBP)** [5] are non-parametric descriptors that summarize the local spatial structure of an image. This method labels the pixels with an LBP code. The LBP codes are obtained by thresholding a given pixel in an n × n neighborhood. The corresponding decimal number of the binary number obtained from the n × n window is assumed to be the LBP code. In recent years LBP has been widely employed for face detection and other computer vision problems such as modelling background and detecting moving object [25]. LBP feature based analysis has also been shown efficient for face recognition [26]. They have been used for extracting features from facial images preceded by active shape mode algorithm to align faces [4].

b**. Scale Invariant Feature Transform (SIFT)** [32] extracts key points from a set of stored images in the database. These features are used to recognize new images, depending upon the Euclidean distance between the feature vectors of the stored images and the new images. The SIFT [28] features of an image are highly distinctive and invariant to rotation and scaling. These characteristics allow a feature to correctly match against a large data of features with high probability. Also, SIFT-extracted features are partially invariant to changes in illuminations and 3D camera viewpoint, making them good candidates for emotion recognition from 3D facial images [38].

c. **Curvelet Transform (CT)** is characterized as a multi-scale directional transform that is employed to obtain almost optimal non-adaptive sparse representation of objects [34]. CT divides each layer into round angles, which are further divided into small directions. Every sub-angle has a coefficient matrix, and these coefficients are used as the feature vector of the images. It has been established in literature that CT has better edge representation and directional characteristics than wavelet features. This advantage of Curvelet transform has been identified by authors and employed for human face recognition in [30].

In this study, we use the features extracted from these three feature extraction methods to train a PBL-McRBFN classifier, and compare their performances.

A. Feature Selection Methods

The features extracted from the methods described in Section IIC may not be all useful and/or discriminative in nature to recognize emotions. Therefore, we employ feature selection methods to choose the most discriminating features for emotion recognition from facial expressions. In this section, we briefly explain the various feature selection methods that we use in this study.

a. **Conditional Mutual Information Maximization (CMIM)**: Mutual Information Maximization (MIM) [44] aims at selecting features that maximize the mutual information between the input features and the class labels. The score F to be maximized by MIM is defined using Mutual Information, I for the j-th feature as follows:

$$F_{MIM}(X_j) = I(X_j; C) \qquad (1)$$

where X_j is the j-th feature and C denotes the class variable. As it can be seen from Eq. (1), MIM assumes the features to be independent of one another. However, features are not always independent of each other. Hence, an additional criterion is introduced for each selected feature in the Conditional Mutual Information Maximization (CMIM) [31]. Accordingly, each feature that is selected must have the least conditional mutual information with the already selected set of features. The CMIM score F is given as:

$$F_{CMIM}(X_j) = \min_{i \in S} I(X_j; C \mid X_i) \qquad (2)$$

where S is the set of features already selected. A subset of gabor features selected through CMIM perform better with higher efficiency in face recognition [37].

b. **Minimal Redundancy Maximum Relevance (mRMR)** [48]: In order to select optimal features from a dataset, apart from considering features having maximum relevance, it is imperative to take into account the criterion of minimum redundancy amongst the selected features, as deleting the redundant although relevant features would not significantly affect the respective class-discriminative power. In [38], authors devised an incremental search algorithm which selects the features which maximize the following score function :

$$F_{mRMR}(X_j) = I(X_j;C) - \frac{1}{m-1} \sum_{X_i \in S_{m-1}} I(X_j;X_i) \qquad (3)$$

where S_{m-1} depict the supposition that we already have a set of m-1 features, and hence the task at hand is to select features from the set $\{X - S_{m-1}\}$ which maximizes the score in Eq. (3).

c. **Double Input Symmetrical Relevance(DISR)**: DISR [40] selects a subset of features from a pool of features based on the rationale that the information on the output class returned by a set of variables is higher than the sum of the information of each variable taken individually. It uses a measure of variable complementarity to derive this relationship. While doing so, if the information returned by combining a set of, say, d features is unknown, the DISR chooses the best scoring sets with d-1 features.

For complete details of the information theoretic feature selection methods discussed in this section, one must refer to [16].

A. Projection Based Learning Algorithm of a Meta-cognitive Radial Basis Function Network Classifier (PBL-McRBFN)

In this section, we explain the architecture and learning algorithm of the PBL-McRBFN briefly. Let the features after extraction and selection using the methods discussed in Section IIB and IIC be given by $\left\{\mathbf{x}^1, ..., \mathbf{x}^t, ..., \mathbf{x}^N\right\}$, where $\mathbf{x}^t = [x_1^t, ..., x_i^t, ..., x_m^t]^T \in \Re^m$ are the m features of the t-th sample. Let the number of emotions to be classified be n. Thus, the target classes of the N samples could be one of $\left\{c_1, ..., c_j, ..., c_n\right\}$. Given these class labels, the coded class labels are given by $\left\{\mathbf{y}^1, ..., \mathbf{y}^t, ..., \mathbf{y}^N\right\}$, where $\mathbf{y}^t = \left[y_1^t, ..., y_j^t, ..., y_n^t\right]^T \in \Re^n$ are the coded class labels for the sample t given by:

$$y_j^t = \begin{cases} 1 & if\ c^t = j \\ -1 & otherwise \end{cases} \quad j = 1, ..., n;\ t = 1, ..., N \qquad (4)$$

The objective of the classifier is to learn the functional relationship between the feature subsets $\mathbf{x}^t$ and their respective target class labels $\mathbf{y}^t$ (i.e., $f : \mathbf{x}^t \rightarrow \mathbf{y}^t$).

The PBL-McRBFN uses a cognitive and a meta-cognitive component to learn this functional relationship through a sequential learning algorithm. The cognitive component has a Radial Basis Function (RBF) network that begins with zero hidden neurons and evolves as the projection based learning algorithm learns the functional relationship represented by the samples.

Without loss of generality, let us assume that there are K hidden neurons after the network has represented t-1 samples. Let us assume the sample t in the data set belong to class l. As the t-th sample is presented to the network, the response of the k-th neuron ,in the hidden layer with Gaussian activation function is given by:

$$h_k^t = \exp\left(-\frac{\left\| x^t - \mu_k^l \right\|^2}{(\sigma_k^l)^2} \right) \qquad (5)$$

Where $\boldsymbol{\mu}_k \in \Re^m$ is the m-dimensional center of the k-th neuron, and $\sigma_k \in \Re$ is the width of the k-th hidden neuron. The responses of the output neurons are given by:

$$\hat{y}_j^t = \sum_{k=1}^{K} w_{jk} h_k^t \qquad (6)$$

The meta-cognitive component of the McRBFN has a dynamic model of the cognitive component and uses a self-regulatory learning mechanism to choose suitable strategies for every sample in the training data set to decide *what-to-learn*, *when-to-learn* and *how-to-learn*. The meta-cognitive component continuously monitors the cognitive component through the maximum hinge loss error, the spherical potential and the confidence of the classifier, and the controls the learning ability of the cognitive component through the sample delete strategy, sample reserve strategy and the sample learning strategy.

(i) Sample Deletion Strategy (*What-to-learn*): Samples with knowledge already represented by the network affect the generalization performance of the network and are hence, deleted. If a sample t satisfies the following sample deletion criteria, it is deleted:

$$\hat{c}^t == c^t \ \text{ AND } \ \hat{p}(c^t \mid x^t) \geq \beta_d \qquad (7)$$

where β_d is the deletion threshold, $\hat{c}^t$ is the predicted class label and c^t is the actual class label for t- th sample.

(ii) Sample Learning Strategy (*How-to-learn*) : The meta-cognitive component uses the following strategies to decide if and when a sample is novel enough to be used in the learning process. When a sample is used in the learning, it is learnt through a projection based learning algorithm.

(a) Neuron addition : When a sample satisfies the following condition, it is used to add a neuron to the cognitive component:

$$(\hat{c}^t \neq c^t \text{ OR } E^t \geq \beta_a) \text{ AND } \psi_c(x^t) \leq \beta_c \qquad (8)$$

where β_a is the self-adaptive metacognitive addition threshold, E^t is the maximum hinge loss error $\psi_c(x^t)$ is the spherical potential, and β_c is the knowledge measurement threshold. β_a is adapted in the following manner based upon hinge error loss E^t:

$$\beta_a := \delta\beta_a + (1-\delta)E^t \qquad (9)$$

Here, $\boldsymbol{\delta}$ is a parameter that dictates the rate of self-adaption, and is kept close to 1. When a neuron is added, the parameters of the neuron are initialized according to its significance to the network.

(b) Neuron deletion: When the contribution of an existing neuron to a class l (Eq. (10)) is less than the pruning threshold β_p, the neuron is deleted.

$$r_k^l = \frac{h(x^i, \mu_k^l)}{\max_j h(x^i, \mu_j^l)} \qquad (10)$$

(c) Parameter update: When a sample presented to the McRBFN satisfies the following condition:

$$\hat{c}^t == c^t \text{ AND } E^t \geq \beta_u \qquad (11)$$

it is used to update the parameters of an existing network. Here, β_u is the self-adaptive parameter update threshold. This threshold β_u is self-adapted according to:

$$\beta_u := \delta\beta_u + (1-\delta)E^t \qquad (12)$$

When a sample is used for parameter update, it uses a projection based learning algorithm [27, 28] to update the parameters of the nearest neuron in the same class.

(iii) Sample Reserve Strategy(*When to learn*) : Samples that do not satisfy the above criteria are reserved for future.

For complete details of the monitory signals and the various learning strategies of the PBL-McRBFN, one must refer to [27, 28, 41]. A guideline to selection of the thresholds can also be obtained from [27, 28, 41].

III. PERFORMANCE STUDY

In this section, we study the performances of the various feature extraction and feature selection methods using the PBL-McRBFN classifier to perform emotion recognition from facial images of the JAFFE data set. The main objective of this study is to choose the best feature extraction and feature selection methods to perform subject independent emotion recognition. Therefore, we study the performances of these methods in their ability to identify discriminative features that address the interpersonal variability in facial expressions.

As seen in Section IIA, The JAFFE dataset has 213 facial images posed by 10 persons, expressing 7 different emotions. In our study, we conduct a 10-fold cross validation with each subject as the testing test in each cross validation. The final accuracy is obtained by averaging over the test accuracy of each fold of cross validation. First, we compare the performance of PBL-McRBFN classifier in emotion recognition using the features extracted using the feature extraction methods discussed in Section IIA, namely, LBP, SIFT and curvelet transform. Next, we use the various feature selection methods discussed in Section IIC on these features and compare the performances of the PBL-McRBFN classifier with the different subsets of features. Finally, we summarize the observations from this study.

A. Performance Comparison of Various Feature Extraction Methods

In this section, we compare the average classification accuracies of the PBL-McRBFN trained using the various feature extraction methods, using the 10-fold subject independent cross validation study.

Table 1: Evaluation of Feature Extraction Methods

Feature Extraction Method	***Average Training Accuracy***	***Average Test Accuracy***
SIFT	98.490±0.99 %	**71.665**±11.49 %
LBP	97.792±3.25 %	61.784±14.61 %
CURVELET	98.948±1.49 %	58.333±17.64 %
PIXEL	98.997±0.62 %	70.705±10.59 %

The 10-fold subject independent cross validation study is conducted by using the facial expression of 9 persons in training and that of the remaining person in testing in each validation. The performances of the PBL-McRBFN classifier trained with features extracted through the LBP, SIFT and curvelet transforms are compared against that of the features as raw pixels in *Table 1*. From the results in Table 1, it can be seen that the PBL-McRBFN trained with the SIFT features outperforms that trained with other feature extraction methods. This could be attributed to its invariability to illumination and 3-D camera view point.

B. Performance Comparison of Various Feature Selection Methods

In this section, we evaluate the performance of PBL-McRBFN with the various features extracted in Section IIIA, selected through the feature selection methods discussed in Section IIC.

Table 2: Evaluation of Feature Extraction Vs Feature Selection Methods

Feature Extraction	***Feature Selection Methods***			
	Complete feature set	***DISR***	***CMIM***	***mRMR***
SIFT	71.665± 11.49 %	76.77± 15.10 %	77.165± 15.62 %	**79.169±** 14.35 %
LBP	61.784± 14.61 %	72.088± 13.08 %	70.048± 12.10 %	**74.642±** 10.08 %
CURVELET	58.33± 17.64 %	73.241± 9.41 %	70.594± 7.78 %	**74.64±** 11.52 %
PIXEL	70.705± 10.59 %	**77.94±** 12.14 %	76.93± 9.97 %	76.807± 10.02 %

Table 2 presents subject-independent average classification test accuracy of the PBL-McRBFN trained with feature subsets selected using various feature selection methods. Performance results show that the subset of features selected through the feature selection methods are more discriminative of the emotions, compared to the entire set of features extracted from the SIFT, LBP, curvelet transform and the pixel based method. The improvement in performance varies between ~5% and ~15%. Moreover, it can also be seen that the PBL-McRBFN trained with the features selected using the mRMR outperformed that with the features selected using the DISR and the CMIM by at least 2%. Thus, it can be observed that the feature subset selected using mRMR from the features extracted with the SIFT method are more discriminative of the emotions from facial expressions.

Next, we conduct a paired t-test [42] to statistically compare the feature extraction and feature selection techniques. Given performances of algorithms A, B and C over multiple datasets, the *paired t-tests* return the probability of events where null hypothesis can be retained or rejected, for each distinct pair of algorithm. The null hypothesis assumes that there is no statistical significance between algorithms. The *p* values from the statistical test signify the confidence of rejecting the null hypothesis. The following observations are made from the results of the paired t-test :

i. mRMR rejects the null-hypothesis and is statistically significant than DISR by $p=0.85$
ii. mRMR rejects the null-hypothesis and is statistically significant than CMIM by $p=0.90$
iii. DISR rejects the null-hypothesis and is statistically significant than CMIM by $p=0.90$

The observations reject the null hypothesis. The observation (i) indicates that the null hypothesis for difference between mRMR and DISR can be rejected with a confidence of $p=0.85$. Thus, although the confidence of rejection of the null hypothesis is not 0.95 to establish a significant difference in performance of the methods, it can be established that the mRMR method of feature selection is different and better than the other feature selection methods. This can also be inferred from the comparison of the classifier with various combinations of feature extraction and feature selection methods in Table 2. Thus, it can be inferred that the mRMR feature selection method on the features extracted from the SIFT method are the most discriminative features for human emotion recognition with a testing accuracy of 79.169.

Next, we compare the performance results of the PBL-McRBFN classifier with features extracted through SIFT and selected with mRMR (SIFT+mRMR+PBL-McRBFFN) with the existing results using state-of-the-art methods in the literature.

A. *Performance Comparison with Existing Results in the Literature*

Several researchers have used several methods to solve the problem of emotion classification using the JAFFE data set. Studies can be categorized into Subject independent study [15, 47] and 10-fold study [45]. In subject independent study aimed at resolving the interpersonal variability in facial expressions, studies have been conducted to recognize one emotion [47] or all emotions [15]. In this section, we compare the results of SIFT+mRMR+PBL-McRBFN with the results reported in these studies. The SIFT+mRMR+PBL-McRBFN has been trained with appropriate data sets, depending on the study.

(a) *10-fold Cross Validation Study*:

In this study, the JAFFE data set is divided into training and testing data sets with 90% of the samples in training and 10% of the samples in testing. We generate 10 subsets of the training data set to conduct a cross validation study through random selection. We then perform 10-fold leave-one-out cross validation by using 9 subsets in training and 1 subset in testing.

Earlier, a 10-fold cross validation study has been conducted with the discrete wavelet transform (DWT) facial features extracted using 2-Dimensional Linear Discriminant Analysis (2D-LDA) trained with Support Vector Machines (SVM) [40]. We compare the performance of SIFT+mRMR+PBL-McRBFN with the results of 2D-LDA+SVM on the DWT extracted features for the 10-fold cross validation study in Table 3. From the Table, it can be observed that SIFT+mRMR+PBL-McRBFN outperforms the DWT + 2D-LDA+SVM by at least 3%.

Table 3 : Accuracy comparison for Subject dependent 10-fold cross validation using JAFFE

ALGORITHMS	ACCURACY
DWT + 2D-LDA + SVM [40]	94.13%
SIFT + mRMR + PBL-McRBFN	**97.334 ± 3.908 %**

(b) *Subject Independent Study to Recognize All Emotions*:

This study aims at recognizing all the 7 emotions in the JAFFE data set. Facial expressions of 9 subjects for all the 7 emotions are used in training and the facial expressions of the remaining subject are used in testing. As the JAFFE data set has facial expression of 10 subjects, 10-fold cross validation study is performed with each subject in the testing set and the remaining 9 for training the PBL classifier. The results of the SIFT+mRMR+PBL-McRBFN in this study is compared with that of Convolutional Neural Networks (CNN) [15] and SVM [41] in Table 4. From the Table, it can be observed that SIFT+mRMR+PBL-McRBFN outperforms the emotion recognition abilities of CNN and SVM by approximately 14% and 17%, respectively. It can also be noted that the SIFT+mRMR+PBL-McRBFN has about 3% lesser variance compared to the SVM and CNN classifiers.

Table 4: Performance Comparison with Existing Results: Subject Independent Study to Recognize All Emotions

ALGORITHMS	ACCURACY
SIFT + mRMR + SVM (RBF)	62.952 ± 17.19 %
CNN [15]	65.22 ± 17.60 %
SIFT + mRMR + PBL-McRBFN	**79.169 ± 14.35 %**

(c) Subject Dependent Single Emotion Recognition:
The JAFFE data set has facial expressions of 10 subjects representing 7 different emotions. Each subject has 3 samples per emotion, and there are approximately 21 samples per subject. In this study, we consider one sample of each emotion of each subject for testing and use the remaining samples for training. Thus, the study considers recognizing emotions per person per emotion involving 70 samples in testing per cross-validation and the remaining samples in training. Table 5 presents the result of SIFT+mRMR+PBL-McRBFN in this performance study, in comparison with LBP+SVM [42] and LBP+Meta-cognitive Fuzzy Inference Systems (LBP+McFIS) [42]. From the table, it can be observed that the SIFT+mRMR+PBL-McRBFN outperforms LBP+SVM and LBP+McFIS, by at least 5% in this study.

Table 5: Performance Comparison with Existing Results: Subject Dependent Single Emotion Recognition

ALGORITHMS	ACCURACY
LBP + SVM [42]	86.66 ± 5.01
LBP + McFIS [42]	87.6 ± 5.79
SIFT + mRMR + PBL-McRBFN	**92.72 ± 5.21**

Thus from the Section IIIC, it can be inferred that the PBL-McRBFN classifier trained with features extracted from SIFT, and selected with mRMR outperforms the state-of-the-art results in the literature on the JAFFE data set, on the different modes of study.

IV. CONCLUSIONS

This paper presents a comprehensive study to understand the most discriminative features for subject independent human emotion recognition from facial images. Three commonly used techniques of feature extraction for emotion recognition, namely, the Local Binary Patterns (LBP), the Shift Invariant Feature Transform (SIFT) and the curvelet transform are depicted. The most discriminative features extracted using these techniques are selected using the DISR, CMIM and the mRMR methods. The feature subsets obtained from these methods on the images from the JAFFE data set are used to train a PBL-McRBFN classifier through a 10-fold cross validation study, leaving one subject in the testing set in each cross validation. From the studies, it is observed that the feature subset selected using the mRMR technique on the features extracted from the SIFT technique are the most discriminative features that can be used for emotion recognition from facial images. A statistical t-test on the various combinations of feature selection and feature extraction methods also reinforces this inference. Performances of the SIFT+mRMR+PBL-McRBFN are also compared with the existing results in the literature for the JAFFE data set, and the performance results show the superior emotion recognition ability of the SIFT+mRMR+PBL-McRBFN.

The increased performance is attributed to both the chosen feature set and the classifier. It is evident from earlier studies [33] that SIFT features are independent of the changes in illumination and 3-D view point. mRMR focuses on reducing the redundancy and selecting the most relevant feature set, capable of identifying the most discriminative feature set. Earlier in [28], the PBL-McRBFN has been shown to perform better classification than other state-of-the-art classifiers for subject independent human action recognition. Therefore, it can be observed from the studies that the best performing PBL-McRBFN classifier trained using the features extracted with SIFT and reduced with mRmR improves emotion recognition performance.

References

[1] C. Ginevra, V. D Santiago,C. Antonio, "Recognising Human Emotions from Body Movement and Gesture Dynamic" in ACII '07 2nd international conference on Affective Computing and Intelligent Interaction, pp. 71 -- 82,2007

[2] H. Kun, Y. Dong, T. Ivan, "Speech Emotion Recognition Using Deep Neural Network and Extreme Learning Machine", in Interspeech 2014, Singapore, pp. 223 -- 227, 2014

[3] C. Chibelushi, F. Bourel, "Facial Expression Recognition: A Brief Tutorial Overview", in CVonline: On-Line Compendium of Computer Vision, editor Robert Fisher, 2003

[4] W. Xun, L. Xingang, L. Lingyun,S. Zhixin ,"A New Facial Expression Recognition Method Based on Geometric Alignment and LBP Features",in 17th Intl. Conference on Computational Science and Engineering, pp. 1734--1737, 2014.

[5] T. Ojala, M. Pietikainen, and T. Maenpaa, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp. 971 – 987, 2002.

[6] H. F. Huang and S. C. Tai, “Facial expression recognition using new feature extraction algorithm,” Electronic Letters on Computer Vision and Image Analysis, vol. 11, no. 1, pp. 41 – 54, 2012.

[7] H. Soyel and H. Demirel, “Facial expression recognition based on discriminative scale invariant feature transform,” Electronics Letters, vol. 46, no. 5, pp. 343 – 345, 2010.

[8] A. Saha and Q. M. J. Wu, “Facial expression recognition using curvelet based local binary patterns,” in IEEE International Conference Acoustics Speech and Signal Processing, 2010, pp. 2470 – 2473.

[9] P. E. and E. R. L., What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS). Oxford University Press, 1997.

[10] [10] J. Ou, X. B. Bai, Y. Pei, L. Ma, and W. Liu, “Automatic facial expression recognition using Gabor filter and expression analysis,” in IEEE International Conference on Computer Modeling and Simulation, 2010, pp. 215 – 218.

[11] S. M. Lajevardi and M. Lech, “Averaged Gabor filter features for facial expression recognition,” in Digital Image Computing: Techniques and Applications, 2008, pp. 71 – 76.

[12] Cohen, N. Sebe, A. Garg, L.S. Chen, T.S. Huang, Facial Expression Recognition From Video Sequences: Temporal and Static Modeling, Computer Vision and Image Understanding (CVIU), Vol. 91, No. 1-2, 2003, pp. 160–187.

[13] V. E. Neagoe, A. D. Ciotec, "SubjectIndependent Emotion Recognition from Facial Expressions using a Gabor Feature RBF Neural Classifier Trained with Virtual Samples Generated by Concurrent Self-Organizing Maps" in Proc. 11th WSEAS Int'l. Conf. Signal Processing, Computational Geometry and Artificial Vision (ISCGAV '11), Florence, Italy, pp. 266-271,2011

[14] R. Valenti, N. Sebe, T. Gevers, "Facial Expression Recognition: A Fully Integrated Approach" in 14th Int’l Conf. Image Analysis and Processing Workshops (ICIAPW 2007), Modena, Italy, 10-13 Sept. 2007, pp. 125-130.

[15] N. Victor-Emil, B. Andrei-Petru, S. Nicu, R. Paul,"A Deep Learning Approach for Subject Independent Emotion Recognition from Facial Expressions" in World Scientific and Engineering Academy and Society, pp. 93--98,2013.

[16] G. Brown "A New Perspective for Information Theoretic Feature Selection", in Proceedings of the Twelfth International Conference on Artificial Intelligence Conference on Artificial Intelligence and Statistics, pp. 49--56, 2009.

[17] M. Lyons, J. Budynek, and S. Akamatsu, “Automatic classification of single facial images,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 21, no. 12, pp. 1357 – 1362, 1999.

[18] M. Turk and A. Pentland, “Face recognition using eigenfaces,” in IEEE conf. Computer Vision and Pattern Recognition, 1991, pp. 586 – 591

[19] X. Feng, M. Pietikainen, and T. Hadid, “Facial expression recognition with local binary patterns and linear programming,” Pattern Recognition and Image Analysis, vol. 15, no. 2, pp. 546 – 548, 2005.

[20] M. Bartlett, G. Littlewort, I. Fasel, and R. Movellan, “Real time face detection and facial expression recognition: development and application to human computer interaction,” in Workshop on CVPR for HCI, 2003.

[21] M. Valstar, I. Patras, and M. Pantic, “Facial action unit detection using probabilistic actively learned support vector machines on tracked facial point data,” in IEEE Conf. Computer Vision and Pattern Recognition, 2005, pp. 76 – 84.

[22] T. O. Nelson and L. Narens. Metacognition: Core Readings, chapter Metamemory: A theoritical framework and new findings, pages 9 – 24. Allyn and Bacon: Boston, T. O. nelson (ed.) edition, 1980.

[23] R. Savitha, S. Suresh, and N. Sundararajan. “A meta-cognitive learning algorithm for a fully complex-valued relaxation network,” *Neural Networks*, vol. 32, pp. 209–218, 2012.

[24] R. Savitha, S. Suresh, and N. Sundararajan, “Metacognitive learning in a fully complex-valued radial basis function neural network,” *Neural Computation*, vol. 24, no. 5, pp. 1297 – 1328, 2012.

[25] K. Subramanian, A. K. Das, S. Suresh, and R. Savitha, “A meta-cognitive interval type-2 fuzzy inference system and its projection based learning algorithm,” *Evolving Systems*, vol. 5, no. 4, pp. 219–230, 2014.

[26] K. Subramanian, S. Suresh, and R. Savitha, “A metacognitive complex-valued interval type-2 fuzzy inference system,” *IEEE Transactions on Neural Networks and their Learning Systems*, vol. 25, no. 9, pp. 1659–1672, 2014.

[27] G. S. Babu, S. Suresh,"Sequential Projection-Based Metacognitive Learning in a Radial Basis Function Network for Classification Problems", in IEEE Transactions on Neural Networks and Learning Systems, vol. 24,no. 2, pp. 194--206, Feb. 2013.

[28] R. V. Babu, R.Savitha, S. Suresh, B. Agarwal, "Subject independent human action recognition using spatio-depth information and meta-cognitive RBF network" in Engineering Applications of Artificial Intelligence, vol. 26, no. 9, pp. 2010--2021, Oct. 2013

[29] J. Michael, A. Shigeru, K. Miyuki, and G. Jiro, “Coding facial expressions with gabor wavelets,” in IEEE Intl. Conf. on Automatic Face and *Gesture Recognition*, pp. 200 -- 205,1998.

[30] Heikkilä M, Pietikäinen M "A texture-based method for modeling the background and detecting moving objects." in IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(4):pp.657-662, 2006 .

[31] C. Shan, S. Gong, P. W. McOwan,"Facial expression recognition based on Local Binary Patterns: A Comprehensive study" in Image and Vision Computing,vol. 27,pp. 803--816, 2009.

[32] D. Lowe and B. Vancouver, “Object recognition from local scaleinvariant features,” in IEEE Intl. Conf. Computer Vision, 1999, pp. 1150 – 1157

[33] D. Lowe "Distinctive Image Features from Scale-Invariant Keypoints" in International Journal of Computer Vision, vol. 60, pp. 91--110,2004.

[34] E. Candes and D. Donoho, “Curvelets- a surprisingly effective nonadaptive representation for objects with edges,” in Curves and Surface Fitting: Saint-Malo, A. Cohen, C. Rabut, and L. Schumaker, Eds., 2000, pp. 105 – 120.

[35] H. Hejazi, M. Alhanjouri ,"Face Recognition Using Curvelet Transform" in International Conference on Industrial, Engineering and Other Application of Applied Intelligent Systems, pp. 388--396, 2010.

[36] F. Fleuret, "Fast Binary Feature Selection with Conditional Mutual Information" in Journal of Machine Learning Research,vol. 5,pp. 1531-1555,2004.

[37] L. Shen, L. Bai, "Information Theory for Gabor Feature Selection for Face Recognition" in EURASIP Journal on Applied Signal Processing(2006),pp. 1--11,2006.

[38] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, No. 8, pp.1226-1238, 2005.

[39] W. Ouarda, H. Trichili, A. M. Alimi and B. Solaiman, "Face recognition based on geometric features using Support Vector Machines" in Soft Computing and Pattern Recognition (SoCPaR), 2014 6th International Conference of, Tunis, pp. 89-95,2014.

[40] P. E. Meyer, C. Schretter, G. Bontempi, "Information-Theoretic Feature Selection in Microarray Data Using Variable Complementarity," IEEE Journal of Selected Topics in Signal Processing, vol. 2,No. 2,pp. 261--274,2008.

[41] G. S. Babu and S. Suresh, “Meta-cognitive RBF networks and its projection based learning algorithm for classification problems,” Applied Soft Computing, vol. 13, no. 1, pp. 654–666, 2013.

[42] J. Demsar, "Statistical Comparisons of Classifiers over Multiple Data Sets" in The Journal of Machine Learning, vol. 7, pp. 1--30, 2006.

[43] S.Berretti, B. B. Amor, M. Daoudi, A. D. Bimbo. "3D facial expression recognition using SIFT descriptors of automatically detected keypoints." in Visual Computer, Springer Verlag, 27 (11), pp.1021-1036,2011.

[44] K. Torkkola,"Feature Extraction by Non-Parametric Mutual Information Maximization" in Journal of Machine Learning Research 3(2003),pp. 1415--1438,2003.

[45] F. Y. SHIH, C. Chuang, P.S.P. Wang "Performance Comparisons of facal expression Recognition in JAFFE Database" in Intl. Journal of Pattern Recognition and Artificial Intelligence,vol. 22, pp. 445-459, 2008.

[46] C. Chang and C. Lin, “LIBSVM: A library for support vector machines,” 2003, National Taiwan University, Taiwan, Deptartment of Computer Science and Information Engineering. [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm/

[47] K. Subramaniam, S. Suresh, R.V.Babu "Meta-cognitive neuro-fuzzy inference system for human emotion recognition" in Intl. Joint Conference for Neural Networks(IJCNN), pp. 1-7,2012.