Exploring a novel method for face image gender Classification using Random Forest and comparing with other Machine Learning Techniques

Amjath Fareeth Basha[1] , Gul Shaira Banu Jahangeer [2]

[1] College of Computers and Information Technology
Taif University, Taif , Saudi Arabia.
[2] College of Computers and Information Technology
Taif University, Taif , Saudi Arabia.

Abstract

Gender classification such as classifying human face is not only challenging for computer, but even hard for human in some cases. This Paper use ORL database contain 400 images include both Male and Female Gender. Our experimental results show the superior performance of our approach to the existing gender classifiers. We achieves excellent classification (100%) accuracy using approach (Continuous wavelet Transform and Random Forest) and compared with other classification Technique like Support Vector Machine, linear discriminate analysis , k- nearest neighbor, Fuzzy c – means, Fuzzy c – means.

Keywords: *Face Gender Classification, Feature Selection, Continuous Wavelet Transform (CWT), Random Forest (RF) Support Vector Machine (SVM), Linear discriminant analysis (LDA), K-nearest neighbors (K-NN).*

1. Introduction

Gender classification using facial images has been in the field of research now days and it is quite interesting. Humans are very good in differentiating the gender from facial images. Social Behavior and human interaction is mainly depending upon on the gender of the person with whom he/she they plan to interrelate. Luckily, human being has the unique capacity of classify gender analyzing simply ones face and exclusive character of conveying personality, emotions, age and lot of other vital information. Even if the face of the human is damaged, to find the gender symptoms, we can identify the gender with very high accuracy [1]. More recently automated gender classification from facial images has gained much interest in computer vision, machine language and Image processing community. The Rapid development progress of this gender classification research area is due to the fast growing field of Internet, electronic commerce, electronic banking systems, more human computer interaction, and demographic research, and security and surveillance applications. It can also bump up other important areas like Image /video indexing, retrieval, passive demographic data collection, vision based human monitoring, human robot interaction , face recognition, face detection, age, traditional determinations are some other important application, where gender classification play a major role.

Many types of gender classification methods are available appearance – based method or holistic approach, geometric or feature based approach, hybrid approach [1][4].In appearance based approach, the whole image or specific regions in a face images is used to generate feature vector .Feature based approach need to localize differential components such as eyes, nose, eyebrows etc. in hybrid approach perceives both local features and whole face. Many techniques have been taken to classify facial images based on gender. This paper works out on the particular approach using Continues Wavelet Transform (CWT) and Random Forest (RF) and compared with other classification technique like Support Vector Machine (SVM), linear discriminate analysis, K- nearest neighbor (K-NN), K-mean, Fuzzy c – means for classifying the gender of the facial images.

To analyze all the features describing an image and to detect gender of the images, it is important to extract all the available gender information from the image. It can be helpful to analyze the image at different resolution levels. Wavelet transform is an ideal tool to analyze images of different gender. It discriminates among several spatial orientations and decomposes images into different scale orientations, providing a method for in space scale representation. The general principles of wavelet transforms have been described elsewhere [5]. Wavelet functions[6] can be used to select the important features for gender classification. In this paper Continuous wavelet Transform have been applied to gender images with varying success. Many authors have developed computerized methods to classify gender face images. In our Proposed Method along with Continuous Wavelet Transform, We Classify the gender using Random Forest (RF) .Our technique performs over well in images containing variations in lighting and facial expression,

pose angles, aging effects etc. Moreover it is less time consuming process

In section 1, we introduce the goals of the paper. Section 2 describes the proposed technique. The Feature Selection using CWT is presented in Section 3.Section 4 discusses classification and prediction using Random Forest. Finally, experimental results with discussion and conclusions are given in section 5 and 6 respectively.

2. **Experimental Setup**

In this section, we describe our experimental setup, which explain about facial dataset and the proposed novel technique with raw data, and analyze the informative features in images containing variations in lighting and facial expression, pose angles, aging effects and finding out the wavelet coefficient using Continuous wavelet transform in gender facial images . Continuous wavelet transform performs better in images containing variations in lighting and facial expression, pose angles, aging effects etc. Moreover it is less time consuming process. Classification and Prediction is done with the Random Forest Classifier. It classifies the gender as male and female with various constrains and do better prediction

2.1 **Dataset Description**

The paper uses the image dataset called ORL Database. The ORL database totally consists of 400 gray scale images representing male and female gender. This images contains variations in lighting, facial expressions, pose, angles, age effects information. In this work, we collect 400 face images out of which 350 faces are male and rest 50 images are female.

Fig 1. ORL Database of 400 images

Table 1: Algorithm for Proposed Method

Step 1: Read an image one by one.
Step 2: Convert the image into single dimensional array.
Step 3: Apply the 1-D Continuous wavelet transform (CWT).
Step 4: Take the coefficient of all images, which is consider along with the label (0 for male, and 1 for female)
Step 5: Repeat the Steps 1 through Step 4 for all the images.
Step 6: Train the dataset using Random Forest.
Step 7: Test the images using Random Forest.
Step 8: Calculate the Classification and prediction rate.

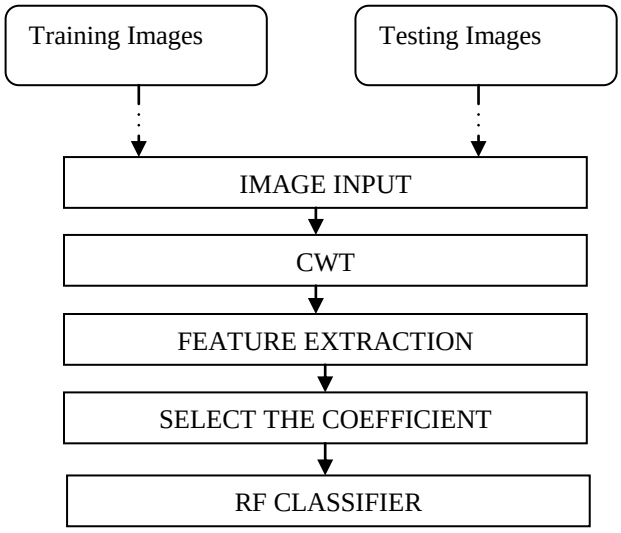

Fig 2. A Block Diagram for the Proposed Methods

3. Feature Extraction:

The feature extraction method that we adopted is Continuous Wavelet Transform (CWT). In This Novel based Method, 1-D continuous wavelet transform allows an input image to be decomposed into a set of independent coefficients corresponding to each one dimensional wavelet basis. We use continuous wavelets to make no redundancy in the information represented by the wavelet coefficients, which leads to efficient representation. Also, it provides exact reconstruction of the original image. Wavelet coefficient [7] represents the "degree of correlation" (or similarity) between the image and the mother wavelet at the particular scale and translation. Thus, the set of all wavelet coefficients [8] gives the wavelet domain representation of the image. After decomposition of the image, the details coefficients can be threshold.

3.1 Continuous Wavelet Transform (CWT):

A continuous wavelet transform (CWT) is used to divide a continuous-time function into wavelet. The CWT has the ability to decompose complex information and patterns into elementary forms. The continuous wavelet transform possesses the ability to construct a time-frequency representation of a signal that offers very good time and frequency localization. The continuous wavelet transform of a continuous, square - integreable function x (t) at a scale a > 0 and translational value b is expressed by the following integral: top of this paragraph illustrates a sub-subheading.

$$X_W(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \psi^* \left(\frac{t-b}{a} \right) dt \quad (1)$$

Where, ψ(t) is a continuous function in both the time domain and the frequency domain called the mother wavelet and represents operation of complex conjugate. The main purpose of the mother wavelet is to provide a source function to generate the daughter wavelets which are simply the translated and scaled versions of the mother wavelet. To recover the original signal x(t), inverse continuous wavelet transform can be exploited.

$$x(t) = \int_0^{\infty} \int_{-\infty}^{\infty} \frac{1}{a^2} (a, b) \frac{1}{\sqrt{|(a)|}} \tilde{\psi} \left(\frac{t-b}{a} \right) db\, da \quad (2)$$

Where, is the dual function of ψ(t). And the dual function should satisfy

$$\int_0^{\infty} \int_{-\infty}^{1} \frac{1}{|a^3|} \psi \left(\frac{t1-b}{a} \right) \tilde{\psi} \left(\frac{t-b}{a} \right) db\, da\, \delta(t - t1) \quad (3)$$

Sometimes ,

$$\tilde{\psi}(t) = C_{\psi}^{-1} \psi(t), where, \quad (4)$$

is called the admissibility constant and is the Fourier transform of ψ. For a successful inverse transform, the admissibility constant has to satisfy the admissibility condition:

$$0 < c_{\psi} < +\infty \quad (5)$$

It is possible to show that the admissibility condition implies that , so that a wavelet must integrate to zero [9].

The advantage of using wavelet-based coding in image compression is that it provides significant improvements in picture quality at higher compression ratios over conventional techniques. Since wavelet transform has the ability to decompose complex information and patterns into elementary forms, it is commonly used in acoustics processing and pattern recognition. Edge and corner detection, partial differential equation solving, transient detection, filter design, Electrocardiogram (ECG) analysis, texture analysis and business information analysis. Continuous Wavelet Transform (CWT) is very efficient in determining the damping ratio of oscillating signals (e.g. identification of damping in dynamical systems). CWT is also very resistant to the noise in the signal.

3.2 Threshold:

For a data set, the mean is the sum of the values divided by the number of values. The standard average, often simply called the "mean". The mean of a set of numbers x_1, x_2... x_n is typically denoted by pronounced "x bar". The mean is often quoted along with the standard deviation: the mean describes the central location of the data, and the standard deviation describes the spread [10]. An alternative measure of dispersion is the mean deviation, equivalent to the average absolute deviation from the mean. It is less sensitive to outliers.

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i \quad (6)$$

The interquartile range (IQR) is a measure of statistical dispersion, being equal to the difference between the third and first quartiles. IQR = Q3 − Q1..The interquartile range of a continuous distribution can be calculated by integrating the probability density function or which yields the cumulative distribution function. The lower quartile, Q1, is a number such that integral of the PDF from -∞ to Q1 equals 0.25, while the upper quartile, Q3, is such a number that the integral from -∞ to Q3 equals 0.75; in terms of the CDF, the quartiles can be defined as

$$Q1 = \text{CDF}^{-1}(0.25),$$
$$Q3 = \text{CDF}^{-1}(0.75), \quad (7)$$

where CDF^{-1} is the quartile function.

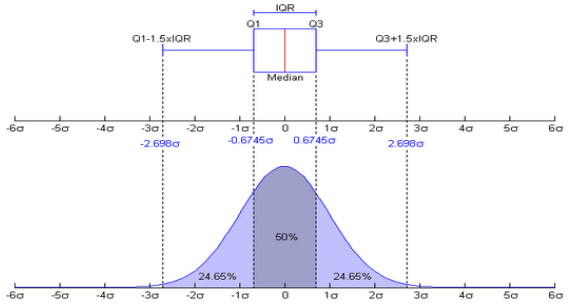

Fig. 3 Diagram of MEAN, IQR, and SIGMA

The IQR is a robust estimate of the spread of the data, since changes in the upper and lower 25 % of the data do not affect it. If there are outliers in the data, then the IQR is more representative than the standard deviation as an estimate of the spread of the body of the data. The IQR is less efficient than the standard deviation as an estimate of the spread when the data is all from the normal distribution

Multiply the IQR by 0.7413 to estimate the second parameter of the normal distribution. based on average threshold of 1 st quartile and 3 rd quartile select the threshold value. The threshold algorithm is shown in table 2.

Table 2: Algorithm for Finding Threshold

Input:
Read N, S from CSV file
N: Coefficients of normal images.
S: Coefficients of sick images.
1) Find mean: N & S.
2) Find IQR: Mean (N & S)
3) Finding Sigma: IQR (N & S)
5) Finding 1st & 3rd quartile of normal and sick.
6) Finding average: avg1, avg2
avg1: Mean (N) +quartile value.
Avg2: Mean(S) + quartile value.
7) Threshold= (avg1+avg2) /2;

4 .Classification

Classification involves segregating the data into segments which are non-overlapping. Any approach to classification assumes some knowledge about the data termed as Training. Training data requires sample input data, domain expertise, and a classification assignment to the data. Performance of classification is measured in terms of classification accuracy.

The outcome of classification can be described as

• **True positive (TP):** A tuple t_i predicted to be in class C_j and is actually in it.
• **False positive (FP):** A tuple t_i predicted to be in class C_j, but is actually not in it.
• **True negative (TN):** A tuple t_i not predicted to be in class C_j, and is actually not in it.
• **False negative (FN):** A tuple ti not predicted to be inclass C_j, but is actually in it.
• **The precision and recall** [11] are used to determine the accuracy of the classifier.

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

A confusion matrix[12] is used to indicate the accuracy. When classifying a class with *m* classes, the confusion matrix is a *m*m* matrix. An entry $C_{i,j}$ indicates the number of tuples assigned to a class C_j but the correct Classification is C_i. A general confusion matrix for two a class is indicated in
Table 3.

Table 3: General Confusion Matrix for Two Classes

Category		Human assignments	
		Yes	No
Classifier assignment	Yes	TP	FP
	No	FN	TN

4.1 Random forest

Random forests (RF) [13] is one of the most successful ensemble learning techniques which have been proven to be very popular and powerful techniques in the pattern recognition and machine learning for high-dimensional classification [13] and skewed problems. These studies used RF to construct a collection of individual decision tree classifiers which utilized the classification and regression trees (CART) algorithms [14].

CART is a rule-based method that generates a binary tree through a binary recursive partitioning process that splits a node based on the yes and no answer of the predictors. The rule generated at each step is to maximize the class purity within the two resulting subsets. Each subset is split further based on the independent rules. CARTs use the Gini index to measures the impurity of a data partition or set of training instances [15]. Although the aim of CART is to maximize the difference of heterogeneity, however, in the real world data sets the over fitting problem that causes the classifier to have a high error of prediction in the unseen data set often encounters. Therefore, the bagging mechanism in RF can enable the algorithm to create classifiers for high dimensional data very quickly [13].

The accuracy of the classification decision is obtained by voting from the individual classifiers in the ensemble. The common element in all of these steps is that the number of *b* tree and a random vector (*Sb*) using bootstrap sample are generated independent of the past random vectors but with the same distribution, and a tree is grown using the training

set and *Sb*. The random forests algorithm is shown in Table 4

Table 4. Algorithm for Random Forest

Input: S: training sample
f: number of input instance to be used at each of the tree
B: number of generated trees in random forest
1) E is empty
2) for b=1 to B
3) Sb = bootstrapSample(S)
4) Cb = BuildRandomTreeClassifiers(Sb,f)
5) E=E ∪{Cb}
6) next b
7) return E

4.2 Support Vector Machine (SVM)

Consider the pattern classifier, which uses a hyper plane to separate two classes of patterns based on given examples {x (i), y (i)} i=1 n .Where (i) is a vector in the input space I=Rk and y (i) denotes the class index taking value 1 or 0.A support vector machine is a machine learning method that classifies binary classes by finding and using a class boundary the hyper plane maximizing the margin in the given training data [16][17][18]. The training data samples along the hyper planes near the class boundary are called support vectors, and the margin is the distance between the support vectors and the class boundary hyper planes. The SVM [16] are based on the concept of decision planes that define decision boundaries. A decision plane is one that separates between assets of objects having different class memberships. SVM is a useful technique for data classification. A classification task usually involves with training and testing data which consists of some data instances. Each instance in the training set contains one "target value" (class labels) and several "attributes" (features).

Given a training set of instance label pairs (xi, yi), i=1… l where xi € R and y € (1,-1)l, the SVM requires the solution of the following optimization problem.

$$\text{Min}_{w,\,b,\,£} 1/2^{w}T^{w} + c\,\Sigma^{l}_{i=1}\,\varepsilon_{I}$$

$$\text{Subject to } y_i\,(w^{T}\,Ø\,(xi) + b) > 1-\varepsilon I,$$

$$\varepsilon_i >= 0$$

Here training vectors xi are mapped into a higher dimensional space by the function Ø. Then SVM finds a linear separating hyper plane with the maximal margin in this higher dimensional space>0.is a penalty parameter of the error term. Furthermore, k (xi, xj) =Ø (xi) Ø (xj) is called the kernel functions.There are number of kernels that can be used in SVM models. These include linear polynomial, RBF and sigmoid.

$Ø = \{x_i * x_j$	linear
$(\gamma x_i x_j + \text{coeff})^d$	polynomial
$\text{Exp}\,(-\gamma\lvert x_i - x_j\rvert^2)$	RBF
$\text{Tanh}\,(\gamma x_i x_j + \text{coeff})$	sigmoid }

The RBF is therefore the most popular choice of kernel types used in SVM. There is a close relationship between SVMs and the Radial Basis Function (RBF) classifiers. In the field of medical imaging the relevant application of SVMs is in breast cancer diagnosis. The SVM is the maximum margin hyper plane that lies in some space. The original SVM is a linear classifier. For SVMs, using the kernel trick makes the maximum margin hyper plane fit in a feature space. The feature space is a nonlinear map from the original input space, usually of much higher dimensionality than the original input space. In this way, nonlinear SVMs can be created. Support vector machinesare an innovative approach to constructing learning machines that minimize the generalization error. They are constructed by locating a set of planes that separate two or more classes of data. By construction of these planes, the SVM discovers the boundaries between the input classes [16]; the elements of the input data that define these boundaries are called support vectors.

For Gaussian radial basis function:

K (x, x') =exp (-|x-x'|2/ (2σ2)).

The kernel is then modified in data dependent way by using the obtained support vectors. The modified kernel is used to get the final classifier.

4.3 Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) [19][20] is used for classification and dimensionality reduction. Linear Discriminant Analysis easily handles the case where the within-class frequencies are unequal and their performances have been examined on randomly generated test data. This method maximizes the ratio of between-class variance to the within-class variance in any particular data set thereby guaranteeing maximal separability. LDA providing better classification compared to Principal Components Analysis. The prime difference between LDA and PCA is that PCA does more of feature classification and LDA does data classification. In PCA, the shape and location of the original data sets changes when transformed to a different space whereas LDA doesn't change the

location but only tries to provide more class separability and draw a decision region between the given classes.
The mathematical operations involved in LDA, the global feature preservation technique is analyzed here. The fundamental operations are:

1. The data sets and the test sets are formulated from the patterns that are to be classified in the original space.

2. The mean of each data set μ_i and the mean of entire data set μ are computed.

$$\mu = \sum_i p_i \mu_i$$

where p_i is priori probabilities of the classes.

3. Within-class scatter S_w and the between-class scatter S_b are computed using:

$$S_w = \sum_j p_j * (cov_j)$$
$$S_b = \sum_j (x_j - \mu)(x_j - \mu)$$

where cov_j the expected covariance of each class is computed as:

$$cov_j = \prod_i (x_j - \mu_i)$$

Note that S_b can be thought of as the covariance of data set whose members are the mean vectors of each class. The optimizing criterion in LDA is calculated as the ratio of between-class scatter to the within-class scatter. The solution obtained by maximizing this criterion defines the axes of the transformed space.
The LDA can be a class dependent or class independent type. The class dependent LDA requires L-class L separate optimizing criterion for each class denoted by C_1, C_2, C_L and that are computed using:

$$C_j = (cov_j)^{-1} S_b$$

The transformation space for LDA, W_{LDA} is found as the Eigen vector matrix of the different criteria defined .

4.4.K-Nearest Neighbors (K-nn)

4.4.1 Knn Classifier:

The simplest classification scheme is a nearest neighbour classification in the image space. Under this scheme an image in the test set is recognized by assigning to it the label of the closest point in the learning set, where distance are measured in image space. If all images have been normalized to be zero mean and have unit variance, then this procedure is equivalent to choosing the image in learning set that best correlates with the test image. Because of normalization process, the result is independent of light source intensity and the effects of a video camera's automatic gain control. Feature selection is achieved using this learning algorithm by constraining each classifier to depend on only a single feature [21].The Euclidean distance metric is often chosen to determine the closeness between the data points in KNN. A distance is assigned between all pixels in a dataset. Distance is defined as the Euclidean distance between two pixels. The Euclidean metric is the function d: Rn X Rn → R that assigns to any two vectors in Euclidean n-space X=(x1,…, xn) and Y=(y1,…, yn) the number.

4.4.2 Knn Algorithm:

1) Each data pixel value within the data set has a class label in theset, Class = {c1,...,cn}.

2) The data points', k-closest neighbors (k being the number ofneighbors) are then found by analyzing the distance matrix.

3) The k-closest data points are then analyzed to determine

which class label is the most common among the set.

4) The most common class label is then assigned to the data point being analyzed..

4.4.3. Knn Performance Vs Choice Of K:

1) When noise is present in the locality of the query instance, the noisy instance(s) win the majority vote, resulting in the incorrect class being predicted. A larger k could solve this problem.
2) When the region defining the class, or fragment of the class, is so small that instances belonging to the class that surrounds the fragment win the majority vote. A smaller k could solve this problem. The KNN shows superior performance for smaller values of k compared to larger values of k. Instances can be considered as points within an n-dimensional instance space where each of the n-dimensions corresponds to one of the n-features that are used to describe an instance. The absolute position of the instances within this space is not as significant as the relative distance between instances. This relative distance is determined by using a distance metric. Ideally, the distance metric must minimize the distance between two similarly classified instances, while maximizing the distance between instances of different classes.
KNN predictions are based on the intuitive assumption that objects close in distance are potentially similar, it makes good sense to discriminate between the k nearest neighbors when making predictions, i.e., let the closest points among the k nearest neighbors have more say in affecting the

outcome of the query point. This procedure has several well known disadvantages. First, if the image in the learning set and test setare gathered under varying lighting conditions, then the corresponding points in the image space will not be tightly clustered. So in order for this method to work reliably under variations in lighting, a learning set which densely sampled the continuum of possible lighting conditions, is required. Second, correlation is computationally expensive. For recognition, we must correlate the image of the test face with each image in the learning set to reduce computational time. Third, it requires large amounts of storage: i.e, the learning set must contain numerous images of each person.

$$d(x,y) = \sqrt{((x1 - y1)2 + ... + (x_n - y_n) 2)}$$

This gives the "standard" distance between any two vectors in R_n.From these distances, a distance matrix is constructed betweenall possible pairings of points (x, y).

4.5 K-Mean:

Clustering algorithms can be applied to solve the segmentation problem. They consist in choosing an initial pixel or region that belongs to one object of interest, followed by an interactive process of neighborhoods analysis, deciding if whether each neighboring pixels belongs or not to the same objects. In this works we use the K- means to resolve the mass detection task on mammograms using texture information obtained from Haralick's descriptors. The K-means algorithm is one of the simplest non-supervised learning algorithms class that solves the clustering segmentation problem [22].

The method follows the usual steps to satisfy the primary objective: clustering all the image objects into K distinct groups. First, K centroids are defined, one for each group, being their initial position very important to the result. After that, it is determined a property region for each centroid, which groups a set of similar objects. The interactive stage of the algorithm is started, in which the centroid of each group is recalculated in order to minimize the objective function. This function, for K-means, is the minimum square method, calculated by Where is the distance metric from any point to the group centroid . Thus, the J (objective function) represents the similarity measure of the n objects contained in their respective groups.

4.6 Fuzzy c – means:

The Fuzzy C-means algorithm [23], also known as fuzzy ISODATA, is one of the most frequently used methods in pattern recognition. Fuzzy C-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. It is based on the minimization of objective function to achieve a good classification. J" is a squared error clustering criterion, and solutions of minimization are least squared error stationary point of "j" .

$$j_m = \sum_{i=1}^{k}\sum_{j=1}^{c} u_{ij\|x_i - c_j\|^2}$$

Where 'm' is any real number greater than 1, is the degree of membership of in the cluster 'j', is the d-dimensional measured data, is the dimension center of the cluster and is any norm expressing the similarity between any measured data and the center. Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, with the update of member ship and the cluster centers.

$$u_{ij} = \frac{1}{\sum_{k=1}^{c}\left[\frac{\|x_i - c_j\|}{\|x_i - c_k\|}\right]^{\frac{2}{m-1}}}$$

The iteration will stop when

$$max_{ij=} \langle |u_{ij}^{k+1} - u_{ij}^{k}| \rangle < \epsilon$$

Where $\in$ is the termination criterion between 0 & 1,whereas k is the iteration steps. This procedure converges to a local minimum or a saddle point of j_m.

5. RESULTS AND DISCUSSION

The experiment is carried out with ORL database containing 400 images of male and female .This images are frontal with variation in pose , expression , and various illumines condition .

In this paper we use feature extraction method CWT on the face gender image and classification method RF. First we compute and select the limited variance of CWT, Coefficients and feed them as inputs to RF. Here we test with 200 images, the no of variable is split in two male gender and female gender and achieve superior classification rate of 100%.

Table 5 : Classification rate and Confusion Matrix of Random Forest

```
  Number of Trees: 10
  No. of Variables tried at each split: 2
  OOB estimate error rate for test data: 0.0000%

        Confusion Matrix For Test Set
       |      1      2 |   err %
  -----|---------------|---------
     1 |    132      0 |   0.0000
     2 |      0     68 |   0.0000


ans =

Percentage of Correct Classification :


per =

   100
```

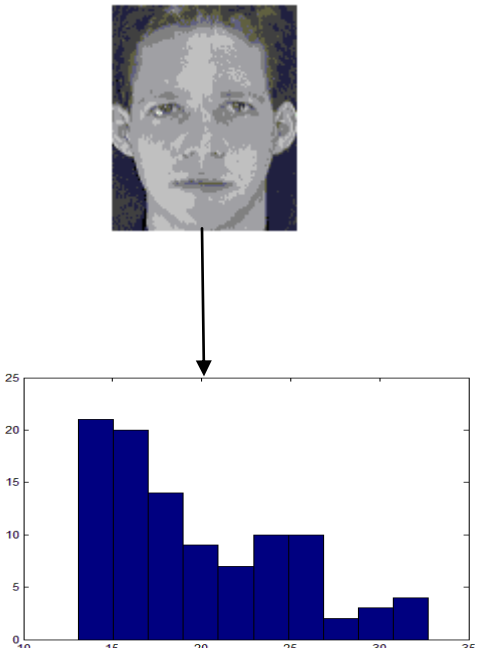

Fig 4. Male Gender Coefficient of CWT

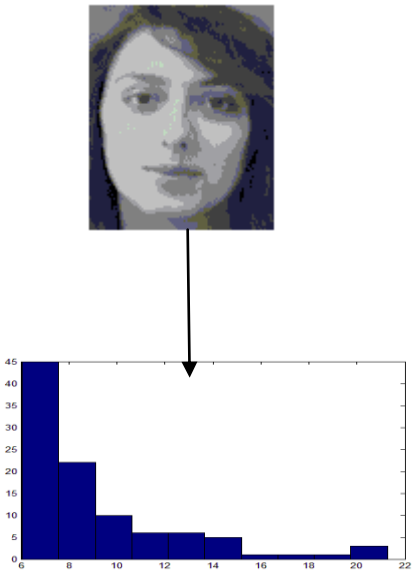

Fig 5. Female Gender Coefficient of CWT

Then we compare our approach with CWT+SVM along with Various Kernels, CWT+ LDA,CWT+ K-mean, CWT+K-NN, CWT+ Fuzzy c – means and the results of error rate are shown in Table 5, 6 and 7. In Figure 4 shows the Male Gender coefficient of continuous wavelet Transform .Similarly Figure 5 shows the Female Gender Coefficient continuous wavelet Transform.

Table 6: Error and Classification rate of CWT and SVM with various Kernels .

CWT +SVM	Error Rate			
Classifier	**Over All**	**Male**	**Female**	**Classification Rate**
svm w/Linear Kernel	2	0	8	98
svm w/RBF Kernel	18	0	72	82
svm w/Poly .Kernel	25	0	100	75
svm w/Quardratic Kernel	22.5	0	90	77.5
svm w/MLP Kernel	14	11.33	22	86

Table 7: Error and Classification rate of CWT w/RF, CWT w/ LDA, CWT w/ K-NN, CWT – K-MEAN, CWT+ Fuzzy c – means.

Classifier	**Error Rate**	**Correct rate**	**Classification Rate**
CWT+RF	0	100	100
CWT+LDA	0.2222	0.778	77.77
CWT+K-NN	25	75	75
CWT – K-MEAN	0.444	0.556	56
CWT+ Fuzzy c – means	0.574	0.426	43

In Table 6 shows the error rate and classification rate of CWT and SVM with various kernels in which SVM using Linear Kernal is more superior when compare to other kernals in SVM. In Table 7 shows the error rate and classification rate of CWT w/RF, CWT w/ LDA, CWT w/ K-NN, CWT – K-MEAN, CWT+ Fuzzy c – means .

We achieves excellent classification (100%) accuracy using approach (Continuous wavelet Transform and Random Forest) and compared with other classification Technique like Support Vector Machine, linear discriminate analysis , k- nearest neighbor, K-Mean,
Fuzzy c – means.

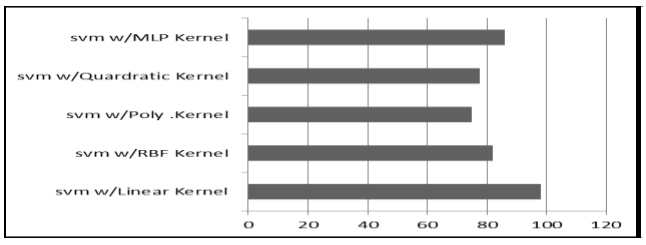

Fig 6: Prediction rate of CWT and SVM with various Kernels.

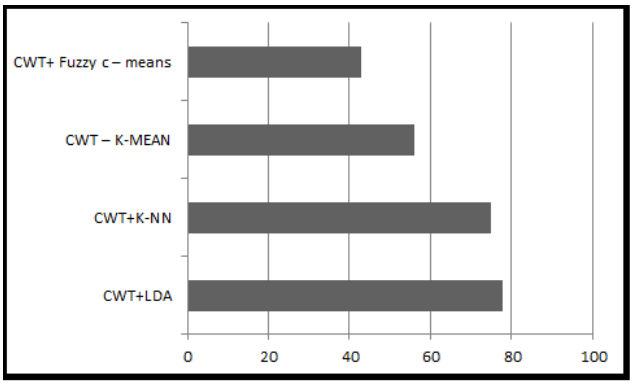

Fig 7: Prediction rate of CWT w/RF, CWT w/ LDA, CWT w/ K-NN, CWT – K-MEAN, CWT+ Fuzzy c – means

In Figure 6 shows the overall error rate of continuous transform and support vector machine with different types of kernels. Similarly Figure 7 shows the prediction rate of various Transform and Support vector Machine with various Kernels.

Table 8. Comparison between the existing methods and the proposed.

Authors	Methods	Percentage
Tamura et al. (1996) [1].	neural networks	93%
Wiskottet al. (1997) [1].	genetic algorithms(GA)	95.3%.
Costen et al. (2004)[1].	SVM	94.42%
Sun et al. (2006)[1].	(SOM)	95.75%,
Lian and Lu (2006)[1].	SVM	96.75%
Baluja and Rowley (2007)[1].	Adaboost classifier	93%
The Proposed Method	CWT (1-D), RF	100%

In Table 8 we present the comparison of techniques, authors to achieve the same goal as this paper. The classification rate of the various techniques are Tamura et al. (1996) is 93 % Using neural networks ,Wiskott et al(1997)is 95.3% using genetic algorithms, Jain and Huang shows the percentage of 99.3% using ICA/LDA. Costene et al.(2002) shows the percentage of 94.42% using SVM, Sun et al.(2006) shows the percentage of 95.75 in SOM. Lian and Lu (2006) show the percentage of 96.75 Using SVM. Baluja and Rowley (2007) show the percentage of 93% using Adaboost Classifier.

Here all the experiments are carried out with grey scale, low resolution images unless otherwise specified with fivefold cross validation. We concluded in our paper RF with CWT, which have no error rate and high percent classification accuracy (100 %), followed by CWT with SVM Linear Kernel which shows the better result and less error rate. More over CWT with SVM (MLP Kernel) also shows the better result. Note that error rate occur more in female compare to male, this is due to less number of female images taken into training set.
The classification rate of our technique is superior to the rest up to techniques. Furthermore 0% error rate in our proposed technique. Overall speaking, the proposed novel technique outperforms other techniques in terms of specificity and sensitivity.

6. Conclusions

An original analysis of algorithm to classifying the gender of male and female is distinguish and computed and verified. The images contain variations in lighting and facial expression, pose angles, aging effects and finding out the wavelet coefficient using Continuous wavelet transform in gender facial images. CWT is constructed to define the feature of the face gender and best and RF is developed for classification. Where we find RF is the best choice for our proposed method.

After Applying CWT, The resulted Coefficient is the binary data: 0 for Male Gender and 1 for Female Gender. The RF classifier with linear kernel exhibits superior efficiency. These algorithms were all coded in MATLAB. Computation time is very less for CWT w/RF when compared with other techniques likes **CWT w/ SVM ,CWT w/ LDA, CWT w/ K-NN, CWT – K-MEAN, CWT+ Fuzzy c – means** .Computational time depends on the configuration of the PC we used. Therefore, it is better to take the ratio of the computational times into account rather than exact values. Finally we concluded that classification rate is raised up to 100% and there is 0% error rate .if CWT and RF are both employed. Finally, RF shows the lowest MSE in Classification and Prediction.

References

[1] Amjath Fareeth and Gul shaira banu Face Gender Image Classification Using Various Wavelet Transform and Support Vector Machine with variousKernels.In

IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 6, No 2, November 2012 .150-157.

[2] Jain, A., Huang, J., May 2004. Integrating independent components and linear discriminate analysis for gender classification. In: Proc. Internat. Conf. on Automatic Face and Gesture Recognition (FGR'04), pp. 159–163.

[3] Saatci, Y., Town, C., April 2006. Cascaded classification of gender and facial expression using active appearance models. In: Proc. 7th Internat. Conf. on Automatic Face and Gesture Recognition (FGR'06), pp. 393–400.

[4] Erno Mäkinen, Roope Raisamo. An experimental comparison of gender classification methods. Pattern Recognition Letters 29 (2008) 1544–1556.

[5] I. Daubechies, "Ten Lectures on Wavelet, "SIAM,pp. 167-213, Philadelphia 1992.

[6].Chen HW. Image processing of micro-calcifications for early-stage breast cancer via wavelet analysis and neural network. M.S. Thesis. Department of Mechanical Engineering, National Cheng-Kung University, Tainan, Taiwan, ROC; 2008.

[7]en.wikipedia.org/wiki/Wavelet.

[8].en.wikipedia.org/wiki/Wavelet transform.

[9].en.wikipedia.org/wiki/Continuous_wavelet_transform.

[10].http://www.mathworks.com/help/toolbox/stats/iqr.html.

[11] http://en.wikipedia.org/wiki/Accuracy_and_precision.

[12] http://en.wikipedia.org/wiki/Confusion_matrix.

[13] Leo Breiman. "Random Forests-Random Features, Technical Report 567, Department of statistics, University Of California, Berkeley, September 1999.

[14] 2001, Breiman L, Random Forests. Machine Learning,45 (1), pp 5-32.

[15] 1984,Breiman L,Friedman J, Olshen R,Stone C,Classification and Regression Trees;Chapman & Hall ; New York.

[16].en.wikipedia.org/wiki/Support Vector Machine.

[17] Y. Liu, Y. F. Zhung, "FS_SFS: A novel feature selection method for support vector machines", pattern recognition New York, vol.39, pg.1333-1345, December 2006.

[18].Smola A. J., Scholkopf B., and Muller K. R., "The connection between regularization operators and support vector kernels", Neural Networks New York, vol.11, pg 637-649, November 1998.

[19] Samarasena Buchala, Neil Davey, Ray J.Frank, Tim M.Gale., "Dimensionality Reduction of Face Images for Gender Classification".

[20] Fahim Mannan, 260 266 294, School of Computer Science, McGill University" Classification of Face Images Based on Gender using Dimensionality Reduction Techniques and SVM.".

[21] Mohammad Kabir Hossain, Abu Ahmed Sayeem Reaz, Rajibul Alam, Dr.William Perrizo, Automatic Face Recognition System using P-tree and K-Nearest Neighbor Classifier.

[22] Leonardo de Oliveira Martins, Geraldo Braz Junior, Arist´ofanes Corrˆea Silva, Detection of Masses in Digital Mammograms using K-means and Support Vector Machine Electronic Letters on Computer Vision and Image Analysis 8(2):39-50, 2009.

[23] Nalini Singh, and Ambarish G Mohapatra ,Silicon Institute of Technology, Bhubaneswar, India, Breast Cancer Mass Detection in Mammograms using K-means and Fuzzy C-means Clustering.