



ISSN: 0067-2904

Machine Learning Approach for Facial Image Detection System

Hind Moutaz Al-Dabbas^{1*}, Raghad Abdulaali Azeez², Akbas Ezaldeen Ali³

¹ Department of Computer Science, College of Education for Pure Science/Ibn Al-Haitham, University of Baghdad, Baghdad, Iraq

² Information Technology Unit, College of Education for Human Science-Ibn-Rushed, University of Baghdad, Baghdad, Iraq

³ Department of Computer Science, University of Technology, Baghdad, Iraq

Received: 5/11/2022 Accepted: 25/12/2022 Published: 30/10/2023

Abstract:

Face detection systems are based on the assumption that each individual has a unique face structure and that computerized face matching is possible using facial symmetry. Face recognition technology has been employed for security purposes in many organizations and businesses throughout the world. This research examines the classifications in machine learning approaches using feature extraction for the facial image detection system. Due to its high level of accuracy and speed, the Viola-Jones method is utilized for facial detection using the MUCT database. The LDA feature extraction method is applied as an input to three algorithms of machine learning approaches, which are the J48, OneR, and JRip classifiers. The experiment's result indicates that the J48 classifier with LDA achieves the highest performance with 96.0001% accuracy.

Keywords: face detection, LDA, J48, OneR, JRip.

نهج التعلم الآلي لنظام اكتشاف صورة الوجه

هند معتز الدباس^{1*}، رغد عبد العالي عزيز²، اقباس عز الدين علي³

¹ قسم علوم الحاسبات، كلية التربية للعلوم الصرفة/ ابن الهيثم، جامعة بغداد، بغداد، العراق

² وحدة تكنولوجيا المعلومات، كلية التربية ابن رشد للعلوم الانسانية، جامعة بغداد، بغداد، العراق

³ قسم علوم الحاسوب، الجامعة التكنولوجية، بغداد، العراق

الخلاصة:

تعتبر فكرة بناء أنظمة اكتشاف الوجه على أن لكل شخص بنية وجه معينة ، وأن مطابقة الوجه المحوسبة ممكنة باستعمال تناظر الوجه. حيث تم استعمال تقنية التعرف على الوجوه لأغراض أمنية في العديد من المؤسسات والشركات في جميع أنحاء العالم. اعد هذا البحث لدراسة التصنيفات في مناهج التعلم الآلي باستعمال استخراج الميزات لنظام الكشف عن صورة الوجه. استعملت طريقة فايولا-جونز لاكتشاف الوجه باستعمال بيانات كبيرة. يتم تطبيق طريقة استخراج الصفات LDA كمدخل لثلاث خوارزميات من مناهج التعلم الآلي وهي:

مصنفات J48 و OneR و JRip. تشير نتائج التجارب إلى أن المصنف J48 مع LDA يحقق أعلى أداء مع دقة 96.0001%.

1. Introduction

In recent years, image classification has gained the interest of many academics due to its vast range of practical applications, including person identification, facial recognition, object categorization, and the diagnosis of disorders in medicine [1]. Due to its use in security systems, video monitoring, and commercial settings, it has become a vital tool for human-computer interaction [2]. Face recognition is the method of recognizing a person's face using a visual system. Due to its intrusive nature and the fact that it is the primary method for identifying humans, face recognition has drawn attention in the wake of artificial intelligence's rapid advancement. Face recognition can be easily checked in an uncontrolled environment without the knowledge of the subject [3 and 4]. Due to the growing number of institutional, military, and commercial applications, face recognition systems make a good topic. Such systems must function with extreme precision and accuracy in order to be trusted. [5 and 6].

Many applications use face detection. Examples of applications that use face detection include Human Computer Interaction (HCI), computer surveillance, biometrics, web search engines, and digital video indexing [7]. The biggest challenges to face detection are pose, illumination, occlusion, and facial expression. Face detection applications have many real-time uses that require high computation time, which is considered to be tough [8 and 9]. Several researchers have implemented face detection with different feature extraction and classification methods. Some have proposed the K-Nearest Neighbor (KNN) algorithm for recognizing faces, depending on appearance-based features that focus on part-based face recognition or on the entire face image [10, 11, and 12]. Other scientists have proposed a facial recognition system using various classifiers, such as Multi-Scale Local Mapped Pattern (MSLMP) [13]. Also used is the non-negative collaborative representation-based classifier (NCRC) [14].

The context of 2D global facial recognition, or using Gabor Wavelet's feature extraction algorithms, feature fusion extracted from histograms of oriented gradient (HOG) and global image descriptor (GIST), as well as detecting the accuracy of the face by using the Viola and Jones algorithm [15, 16, and 17]. However, the Viola-Jones method is used by several scientists for face image detection and has been implemented in a variety of applications because of its high degree of accuracy and speed. Accordingly, using the Viola-Jones algorithm with the LDA method is considered the novelty of this research. These features will be the inputs of the machine learning classifier model to achieve the highest accuracy.

The aims of this work are to propose three algorithms to examine the classifications in machine learning approaches with feature extraction for the facial image detection system, as well as implement the huge MUCT database and compare the results with other researchers who used the same databases depending on performance measures.

2. Literature Survey

Here is a survey of some researchers who have implemented face detection with different feature extraction and classification methods.

Setiawan and Muttaqin [10] proposed the K-Nearest Neighbor (KNN) algorithm for recognizing faces on an ARM processor. To reach the best k-value to create proper face recognition with a low-power processor, use principal component analysis (PCA) and linear discriminant analysis (LDA) for feature extraction.

Barnouti et al. [11] proposed face recognition system depends on appearance-based features that focus on the entire face image rather than local facial features. The Viola-Jones face detection method was used. Feature extraction and dimension reduction methods were applied, using principal component analysis (PCA) and linear discriminant analysis (LDA). Square Euclidean Distance (SED) is used to measure the distance between two images.

Silva et al. [12] proposed a facial recognition system based on multi-scale local mapped patterns (MSLMP), using genetic algorithms (GA) to optimize parameters and weighting matrices. To deal with difficult databases like MUCT, this technique was based on the average gray levels of the images in the database. The results obtained for the database are superior and have high accuracy.

Zhou and Zhang [13] proposed a collaborative representation by directly using non-negative representations to show a test sample collaboratively, termed a non-negative collaborative representation-based classifier (NCRC). A rectified linear unit (ReLU) was used to collect all NCRC functions that perform filtering on the coefficients obtained by l_2 minimization according to the collaborative representation-based classification (CRC) objective function. Experiments on four different databases, including face and palm prints, revealed promising efficiency and accuracy results for the method.

Moreano and Palomino [14] proposed the context of 2D global facial recognition using Gabor Wavelet feature extraction algorithms and facial recognition Support Vector Machines (SVM) with linear, cubic, and Gaussian kernel functions. The models generated by the technique allow for the execution of tests with high accuracy in facial recognition.

Color Component Selection and Color Component Fusion were proposed by Nguyen-Quoc and Hoang [15] as methods for extracting HOG from color images, as well as extended kernels to improve HOG performance. With our new approaches in color component analysis, the experimental results of several facial benchmark datasets are enhanced with high accuracy.

Rashed and Hamd [16] proposed an automatic face recognition system based on features focusing on the whole image as well as local-based features using the Local Binary Pattern Histogram (LBPH), Principal Component Analysis (PCA), and Linear Discriminant Analysis (LDA). In addition, the system generated the machine learning algorithms PART and J48. The results showed high accuracy for detection and feature extraction.

Szmurło and Osowski [17] proposed an ensemble of classifiers based on the CNN architecture for the problem of increasing the generalization ability of classification systems and compared different structures of the ensemble. The results are combined with three classifiers: softmax, support vector machine (SVM), and random forest of the decision tree. The results of experiments have shown a high improvement in class recognition resulting from the application of a properly designed ensemble.

3. Methodology

One of the visual tasks that humans can perform with ease is face detection, but in computer vision, this task is regarded as challenging. There are 3755 images of faces and 76 manual landmarks in the Milborrow/University of Cape Town (MUCT) database, which is the one that was used. Lighting, age, and ethnicity diversity are all incorporated into the database. All of the photographs in this collection from December 2008 were taken of people near the University of Cape Town campus. Each person in this database university students, parents, high school

professors, and employees—has been captured on camera by five different cameras, making the database ideal for applications that need to see the same person repeatedly [16]. Sample images from the MUCT database are shown in Figure 1 [18].



Figure 1: Sample images of MUCT database.

Every image in the dataset has its facial region detected and cropped using the Viola-Jones method. LDA feature extraction is used for dimension reduction. Machine learning-based classification models are used. Figure 2 shows the system methodology process.

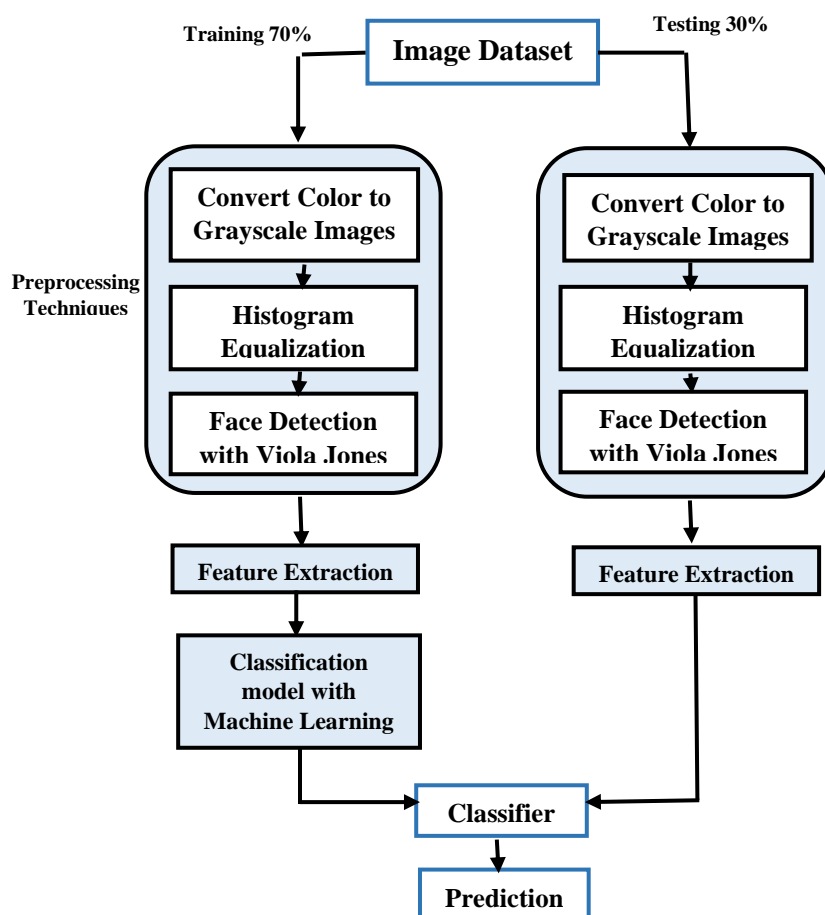


Figure 2: The system methodology process.

3.1 Preprocessing Techniques

A preprocessing stage is implemented as the first phase. It uses many methods to enhance the input images, including converting color images to grayscale images, histogram equalization, detection cropping, and resizing.

3.1.1 Convert Color to Grayscale Images

The image domain is a mathematical model with different color components representing color details, and the determination of the color model is important before the analysis process of the color image begins. The color models provide color data for each pixel in a particular image [19]. A grayscale image's brightness has been represented using an 8-bit value, but a color image's pixel color is represented using a 24-bit value. The brightness of pixels in grayscale images spans from 0 to 255, with 0 intensity denoting black and 255 intensity denoting white [19, 20]. To convert the image into a grayscale, this is brought on by the fact that the three colors: Red (R), Green (G), and Blue (B) should be averaged. Since each of the three hues has a unique wavelength and contributes differently to the creation of an image, the average must be calculated according to each color's contribution rather than just utilizing the average approach. This has been given by the luminosity method. It indicates that the contribution of the color red must be reduced, the contribution of the color green must be increased, and the contribution of the color blue must be placed between these two. Figure 3 illustrates samples of the grayscale of the images [21, 22]. Equation (1) depicts the process of converting a colored image to a grayscale image.

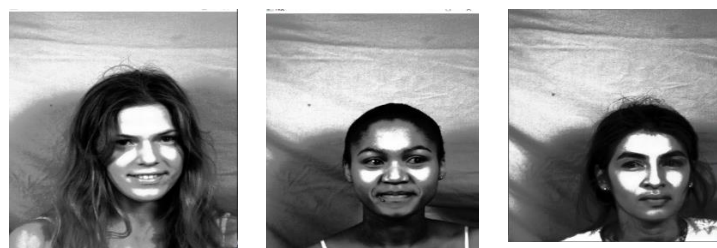
$$\text{New grayscale image} = (0.3 * R) + (0.59 * G) + (0.11 * B) \quad \dots \dots (1)$$



Figure 3: Samples of the grayscale images.

3.1.2 Histogram equalization (HE)

Histogram equalization is a method of computer image processing used to increase contrast in images. In order to improve low-contrast images' image quality and face recognition capabilities, it is common practice to distribute the most frequent intensity values evenly [23]. The image's dynamic range (contrast range) is altered as a result, making some crucial facial features more noticeable [24]. Figure 4 shows a sample of HE images.



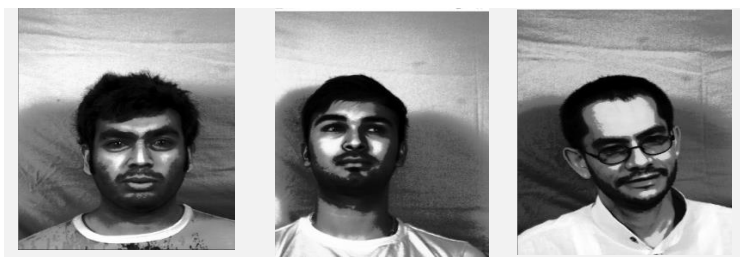


Figure 4: Histogram Equalization images.

3.1.3 Face Detection with Viola Jones Algorithm

The Viola Jones algorithm is a learning-based algorithm that is employed for object detection due to its high degree of accuracy and speed. Figure 5 shows a sample of the detected image after detection with the Viola Jones algorithm. The Viola-Jones method consists of four concepts: Haar features, integral image formation, Adaboost, and cascading [25].



Figure 5: The detected images.

A) Haar Features

The entire image is divided into small windows or rectangular areas of size $M \times M$. Each window's features are determined separately. For face detection, three different feature types are typically used: two rectangles, three rectangles, and four rectangles. The difference in the sums of the pixels within two rectangular sections is known as the "two-rectangle" characteristic. These two rectangular areas are near one another, either horizontally or vertically, and are the same size and shape. A three-rectangles feature adds the sum of the pixels in the middle rectangle to the sums of the pixels in the two outside rectangles. The difference between diagonal pairs of rectangles is computed by a four-rectangle feature, which is the final step [26]. These features are shown in Figure 6.

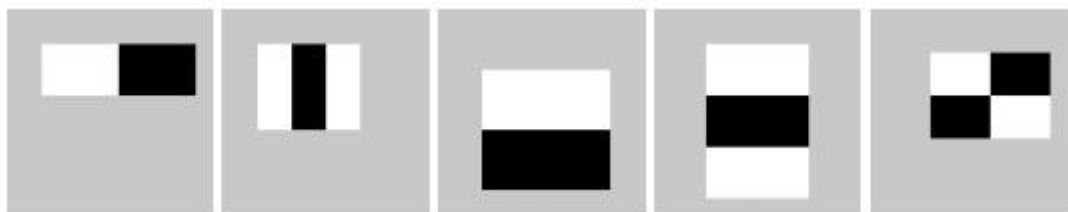


Figure 6: Haar features on the detected face.

B) Integral image formation

An intermediate image representation known as an integral image is used to quickly calculate Haar-like features. The calculation formula of an integral image is displayed in

equation (2). The integral image at location x, y contains the sum of the pixels above and to the left of (x, y) [25].

$$P(x,y)=\sum_{s=1}^x \sum_{t=1}^y I(s, t); 1 \leq x \leq M, 1 \leq y \leq M \quad \dots \dots (2)$$

Where the integral image (P), which is displayed in Figure 7. (I) refers to the original image.

The rectangle's region is represented by (M). The integral image at location 1 in Figure 7 is the sum of pixels in region A. At position 2, the total of the pixels in region A+B is given; at position 3, the total of the pixels in region C+A is given; and at position 4, the total of the pixels in region A+B+C+D is given.[20].

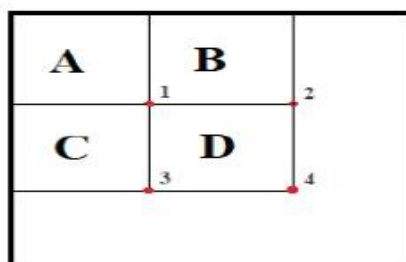


Figure 7: integral of image information.

C) Adaboost algorithm for feature classification

There are a large number of Haar features computed for each window. A majority of these features are unnecessary. Using this approach, the redundancy is reduced. A function for learning classifications is the Adaboost algorithm, which reduces the size of a large set of characteristics by eliminating useless ones. In essence, it is a classifier created by weighing a number of weak classifiers. Each and every window feature is a poor classifier [20]. As it chooses the best features out of all the options, this method may also be thought of as a feature selection algorithm. The features that were chosen are the ones that are most effective in representing a face. This technique reduces thousands of features to a few hundred [26].

D) Cascading

After choosing the finest features from each window, it should maintain which of these windows contain faces. In an image, only 0.01% of the windows are normally positive, suggesting they include people. The first discovered faces must proceed through a number of cascaded phases before finding the positive windows. Every level lowers the quantity of false positives, or areas that are mistakenly identified as faces. A classifier is created for each stage, utilizing a few features. Every level after that adds more and more features, increasing the complexity of the classifier. Every stage has the option of rejecting the discovered region or proceeding to the next. Therefore, only the area that successfully completes all stages is categorized as a face [27]. On each database, the Viola-Jones method was used to detect faces; this technique had a high detection rate. Also, each image was identified, resized using bicubic interpolation, and then cropped to a 100x100 size. Figure 8 shows a sample of the images after resizing and cropping. Bicubic interpolation is an ideal method; it would also be a good option if quality were an issue. In bicubic interpolation, the sixteen nearest neighbors of a pixel have been considered. The intensity value assigned to point (x,y) is obtained using equation (3).

$$v(x, y) = \sum_{i=0}^3 \sum_{j=0}^3 a_{ij} x^i y^j \quad \dots \dots (3)$$



Figure 8: Images after resizing stage.

3.2 Feature Extraction

In order to reduce the complexity of the space and the time required for machine training, the feature extraction phase is crucial to the process of dimensionality reduction. It involves extracting the relevant feature subset from the original data using a series of rules. The input data is transformed into a set of features during feature extraction, and the resultant reduced representation retains the majority of the pertinent data from the original data [28].

3.2.1. Linear Discriminant Analysis (LDA)

LDA is a method for supervised linear dimensionality reduction that seeks the subspace that distinguishes between several classes the most effectively. LDA's main objective is to apply dimensionality reduction while retaining as much class-discriminatory data as is practical. It reduces the dimensions of a high-dimensional feature vector [29]. The goal of discriminant analysis is to classify different types of objects, such as people, clients, goods, etc., into one of two or more sets based on a variety of characteristics, such as weight, age, preference score, gender, and income. A linear discriminant model can be used if one can assume that the sets are linearly separable. According to the idea of linear separability, qualities that identify the items can be grouped into linear groups to split the sets. If there are two features, there will be lines dividing the group of objects. If there are three features, the partition is a plane; if there are more features, such as independent variables, the partition is a hyper-plane. Here are the further LDA steps [30]:

➤ Examples for class-1 and class-2

➤ Compute the mean of class-1 and class-2 as μ_1 and μ_2

➤ Class-1 and class-2 covariance matrices (C_1 and C_2)

➤ use equation (4) to compute within-class scatter matrix

$$S_w = C_1 + C_2 \quad \dots \dots (4)$$

➤ Equation (5) is used to compute the among-class scatter matrix

$$S_B = (\mu_1 - \mu_2) * (\mu_1 - \mu_2) \quad \dots \dots (5)$$

➤ Calculate the mean for all classes.

➤ The general eigenvalue issue is then resolved with the help of the LDA plan, see equation (6) and (7)

$$S_w^{-1} S_B W = \lambda W \quad \dots \dots (6)$$

$$W = \text{eig}(S_w^{-1} S_B) \quad \dots \dots (7)$$

where W is projection vector.

3.3 Classification model with Machine Learning (ML)

Classification is the process of building a model of classes from a set of records with class labels. A group of algorithms known as "machine learning" use a set of data to learn a model.

Machine learning research places a lot of emphasis on developing software that can automatically recognize complicated patterns and draw conclusions from data [31]. J48, OneR, and JRip are used for classification in this study.

2.3.1 J48 decision tree algorithm

The decision tree algorithm is used to determine how the characteristic vector acts in various situations. The classes for the newly produced instances are also being found based on the training instances. This algorithm creates the rules for the target variable's prediction. The data's crucial distribution can be understood with the use of the tree classification technique [32]. The additional properties of J48 are accounting for missing values, decision tree pruning, continuous attribute value ranges, derivation of rules, etc. In the WEKA data mining tool, J48 is an open source Java implementation of Quinlan's C4.5. This algorithm was chosen due to its wide acceptance and use in ML. The WEKA tool provides a number of options associated with tree pruning. In cases of potential overfitting, pruning can be used as a tool for précising. In other algorithms, the classification is performed recursively until every single leaf is pure; that is, the classification of the data should be as perfect as possible. This algorithm generates the rules from which the particular identity of that data is generated. The objective is the progressive generalization of a decision tree until it gains equilibrium between flexibility and accuracy. Algorithm 1 shows the steps of the J48 algorithm.

Algorithm 1: J48 Algorithm steps
Input: Features extracted by LDA
Output: Build a classifier
Processing steps:
Step 1: The leaf is labeled with a similar class if the instances belong to the same class.
Step 2: For each attribute, the potential data will be generated and the gain in the data will be derived from the test on the attribute.
Step 3: The best attribute will be chosen according to the current selection parameter.

3.3.2 One Rule (OneR) algorithm

OneR, which stands for "one rule," is an easy-to-understand, accurate classification method that builds one rule for each predictor in the data before selecting the rule with the smallest overall error as its "one rule" and limiting decision trees to level one. To create a rule for a predictor, it would produce a frequency table for each predictor against the objective. It has been shown that OneR generates rules that are intuitive to people and just slightly less accurate than state-of-the-art classification algorithms [33]. Algorithm 2 shows the steps of the OneR algorithm.

Algorithm 2: OneR Algorithm steps
Input: Features extracted by LDA
Output: Build a classifier
Processing steps:
Step 1: For each predictor,
Step 2: For each value of that predictor, make the following rule;
1. Count how often each value of target class appears
2. Find the most frequent class
3. Make the rule assign this prediction value to that class.
4. Calculate the total error for each predictor's rules.
Step 3: Choose the predictor with the smallest total error.

3.3.3 JRip algorithm

JRip was implemented by Cohen, W. W., in 1995. This algorithm was implemented as a propositional rule learner, Repeated Incremental Pruning to Produce Error Reduction (RIPPER). By altering or updating certain rules, Cohen has implemented RIPPER to improve the accuracy of the rules. Reduce Error Pruning was used to select when to stop adding new conditions to a rule after isolating some data for training [33]. By employing the minimum description length heuristic as a stopping condition. The induction rule's post-processing phases revise the rules for the estimates produced by the global pruning technique, which increases accuracy. Algorithm 3 shows the steps of the JRip algorithm.

Algorithm 3: JRip algorithm steps
Input: Features extracted by LDA
Output: Build a classifier
<p>Processing steps:</p> <p>Step 1: For each class, initialize a set of rules called $RS = \{ \}$,</p> <p>Step 2: Building phase: repeated step 3 and step 4 until reaching a stopping criterion, at which point the entire set of rules are optimized.</p> <p>Step 3: Grow a rule: Grow one rule by greedily adding antecedents (or conditions) to the rule until the rule is perfect. Every conceivable value for each attribute is tested, and the condition with the greatest information gain is chosen by the method.</p> <p>Step 4: Prune a rule: Incrementally prune each rule and allow the pruning of any final sequences of the antecedents.</p> <p>Step 5: Optimization: Optimize the rule set</p>

4. Results and Discussion

The J48, JRip, and OneR classification algorithms were applied to the data. The classification accuracy tests results will be compared among the three algorithms in order to select the best one. Several metrics have been used in the system and are designed for evaluating system performance.

A) Precision: is the amount of the true positives that are separated by total number of true positive cases and the total number of the false positives [16].

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

Where: -

TP: True Positives

FP: False Positives

B) Recall: The ability to find every relevant example in a data set, where "precision" represents the ratio of data points that this model says are relevant that are actually relevant [17, 20].

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

Where: -

TP: True Positives

FN: False Negatives

C) F-measure: The harmonic medium value of accuracy and recall, with F1 being the optimal value at one and worse at zero. [19]

$$F_1 = \frac{1}{\frac{1}{recall} + \frac{1}{precision}} = 2 * \frac{precision * recall}{precision + recall} \dots \dots (10)$$

Table 1 shows the experimental results of each algorithm depending on the performance measures.

Table 1: The experimental results of J48, JRip, and OneR classifications

classification algorithms	Precision	Recall	F-measure
J48	96.0001%	95.4532%	95.3701%
JRip	84.0410%	78.8776%	79.9410%
OneR	61.2693%	73.0935%	65.2130%

The accuracy of these classifications is typically evaluated in order to assess the effectiveness of classification algorithms. It is evident from Table 1 that the implemented J48 with LDA feature extraction has the highest classification accuracy, where the precision is 96.00%, the recall is 95.45%, and the F-measure is 95.37%. The second highest classification accuracy for JRip with the LDA feature extraction algorithm is 84.04%, the recall is 78.87%, and the F-measure is 79.94%. Moreover, the OneR algorithm showed the lowest classification accuracy of other algorithms: the precision is 61.26%, the recall is 73.09%, and the F-measure is 65.21%. So the J48 algorithm outperforms the other state-of-the-art classification methods in terms of classification accuracy. Table 2 shows the comparison between the results obtained by this proposed system and some previous related works that used the MUCT database.

Good accuracy rates (90-93%) were obtained in Setiawan and Muttaqin [10], Silva et al. [12], Moreano and Palomino [14], and Nguyen-Quoc and Hoang [15] using different feature extractions like PCA, LDA, and Gabor wavelets, with different classifiers like KNN, SVM, HOG, and GA. Using Euclidean distance and random forest, Barnouti et al. [11], Szmuro and Osowski [17] obtained good accuracy of 87.5% and 83.3%, respectively. With an accuracy of 77.78%, Zhou and Zhang [15] obtained a good result using novel collaborative representation at the NCRC. Rashed and Hamd [16] used the J48 classifier and reached 94.337%. In this proposed system, the higher accuracy is 96.0001% by using LDA feature extraction and the J48 classifier, so it is considered the higher accuracy for the proposed system.

Table 2: comparison between proposed system and some previous works

Methods of previous works	Feature Extraction	Classifier	Accuracy
Zhou and Zhang [13]	-	SVM, KNN, NCRC	25.56%, 27.87%, 77.78%
Szmurlo and Osowski [17]	-	Random Forest	83.3%
Barnouti et al. [11]	PCA and LDA	Euclidean distance	87.5%
Setiawan and Muttaqin [10]	PCA and LDA	KNN	91.5%
Silva et al. [12]	MSLMP	GA	93.49%
Nguyen-Quoc and Hoang [15]	-	Histogram of Oriented Gradient (HOG)	93.47%
Moreano and Palomino [14]	Gabor Wavelet's	SVM	93.70%
Rashed and Hamd [16]	LDA	J48	94.337%

Proposed work	LDA	J48	96.0001%
---------------	------------	------------	-----------------

5. Conclusions

Three algorithms to examine the classifications in machine learning approaches with feature extraction for the facial image detection system are proposed. Implementing the Milborrow/University of Cape Town (MUCT) database that took into consideration lighting, age, and ethnic diversity. The Viola-Jones method is used to detect and crop the face region in each of the database images. LDA feature extraction is used for dimension reduction. Classification models based on machine learning are utilized. The decision tree algorithm is used to determine how the characteristic vector acts in various situations. Understanding the data's crucial distribution is accomplished with the use of the tree classification technique, which generates rules that are easy to understand. Reduce Error Pruning was used to select when to stop adding new conditions to a rule after isolating data for training. The J48, JRip, and OneR classification algorithms are applied to the data.

The classification accuracy results will be compared with the three algorithms. The implemented J48 with LDA feature extraction has the highest classification accuracy (96.00% precision, 95.45% recall, and 95.37% F-measure). The second highest classification accuracy for JRip with the LDA feature extraction algorithm is 84.04% precision, 78.87% recall, and 79.94% F-measure. The lowest classification accuracy is the OneR algorithm (61.26% precision, 73.09% recall, and 65.21% F-measure). The results of the experiments indicate that the J48 classifier with LDA achieves the best performance with 96.0001% accuracy, which overtakes the other classifiers, OneR and JRip. Comparing the results with other researchers who used the same databases based on different performance measures indicated that the proposed system had a higher accuracy of 96.0001% by using LDA feature extraction and the J48 classifier.

References

- [1] K. H. Teoh et al., "Face Recognition and Identification Using Deep Learning Approach," *Journal of Physics: Conference Series*, vol. 1755, no. 1, pp. 1-9, Feb. 2021.
- [2] R. A. Azeez, "Determination Efficient Classification Algorithm for Credit Card Owners: Comparative Study," *Engineering and Technology Journal*, vol. 39, part B. no. 01, pp. 21-29, 2021.
- [3] N. H. Ali, et al., "Constructed model for micro-content recognition in lip reading based deep learning," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 5, pp. 2557-2565, October 2021.
- [4] S. Karki, et al., "A Novel Solution of an Enhanced Error and Loss Function using Deep Learning for Hypertension Classification in Traditional Medicine," *IEEE, In proceeding of 5th International Conference on Innovative Technologies in Intelligent Systems and Industrial Applications (CITISIA)*, pp. 1-10, 2020.
- [5] F. G. Mohammed and H. M. Al-Dabbas, "The Effect of Wavelet Coefficient Reduction on Image Compression Using DWT and Daubechies Wavelet Transform," *Science International*, vol. 30, no. 5, pp.757-762, 2018.
- [6] N. H. Ali et al., "Learning Evolution: A Survey," *Iraqi Journal of Science*, vol. 62, no. 12, pp. 4978-4987, 2021.
- [7] M. E. Abdulmunim and F. B. Ibrahim, "Design of an Efficient Face Recognition Algorithm based on Hybrid Method of Eigen Faces and Gabor Filter," *Iraqi Journal of Science*, vol. 57, no. 3B, pp. 2102-2110, 2016.
- [8] F. B. Ibrahim and M. E. Abdulmunim, "Real Time Face Recognition System based Hybrid Method," *International Journal of Scientific Engineering and Applied Science (IJSEAS)*, vol. 4, no. 4, May 2018.

- [9] A. H. Morad and H. M. Al-Dabbas, "Classification of Brain Tumor Area for MRI images," *1st International Conference on Pure Science (ISCPS), Journal of Physics*, vol. 1660, pp. 012059, 2020.
- [10] E. Setiawan and A. Muttaqin, "Implementation of K-Nearest Neighbors Face Recognition on Low-power Processor," *Telekomnika*, vol. 13, no. 3, pp. 949-954, 2015.
- [11] N. H. Barnouti et al., "Face Detection and Recognition Using Viola-Jones with PCA-LDA and Square Euclidean Distance," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 7, no. 5, pp. 371-377, 2016.
- [12] E. M. Silva et al., "Face Recognition Using Local Mapped Pattern and Genetic Algorithms," *In proceedings of the International Conference on Pattern Recognition and Artificial IntelligenceAt*, pp. 11-17, 2018.
- [13] J. Zhou and B. Zhang, "Collaborative Representation Using Non-Negative Samples for Image Classification," *Sensors*, vol. 19, no. 11, pp. 1-14, 2019.
- [14] J. Moreano and N. Palomino, "Efficient Technique for Facial Image Recognition with Support Vector Machines in 2D Images with Cross-Validation in Matlab," *Wseas Transactions on Systems and Control*, vol. 15, pp. 175-183, 2020.
- [15] H. Nguyen-Quoc and V. T. Hoang, "A Revisit Histogram of Oriented Descriptor for Facial Color Image Classification Based on Fusion of Color Information," *Journal of Sensors*, vol. 2021, pp. 1-12, Nov. 2021.
- [16] A. H. Rashed and M. H. Hamd, "Robust Detection and Recognition System based on Facial Extraction and Decision Tree," *Journal of Engineering and Sustainable Development*, vol. 25, no. 4, pp. 40-50, 2021.
- [17] R. Szmurło and S. Osowski, "Ensemble of classifiers based on CNN for increasing generalization ability in face image recognition," *Bulletin of the Polish Academy of Sciences: Technical Sciences*, vol. 70, no. 3, pp. 1-10, 2022.
- [18] "The MUCT Face Database," <http://www.milbo.org/muct/>.
- [19] F. G. Mohammed and H. M. Al-Dabbas, "Application of WDR Technique with different Wavelet Codecs for Image Compression," *Iraqi Journal of Science*, vol. 59, no. 4B, pp. 2128-2134, 2018.
- [20] H. Nguyen-Quoc and V. T. Hoang, "Face recognition based on selection approach via Canonical Correlation Analysis feature fusion," *IEEE, Zooming Innovation in Consumer Electronics International Conference (ZINC)*, pp. 54-57, 2020.
- [21] J. M. AL-Tuwaijari et al, "Performance Evaluation of Face Image Recognition Based Voila-Joins with SVM," *Iraqi Journal of Information Technology*, vol. 9, no. 1, pp. 48-63, 2018.
- [22] I. A. Abdulmajeed, "Face Detection and Recognition Based on Skin Color," *M.Sc. thesis*, University of Baghdad, College of Science, Computer Science department, 2021.
- [23] C. Saravanan, "Color Image to Grayscale Image Conversion," *IEEE, Second International Conference on Computer Engineering and Applications*, pp. 196-199, 2010.
- [24] N. Salem et al., "Medical image enhancement based on histogram algorithms," *Procedia Computer Science*, vol. 163, pp. 300-311, 2019.
- [25] M. K. Dabhi and B. K. Panchol, "Face Detection System Based on Viola - Jones Algorithm," *International Journal of Science and Research (IJSR)*, vol. 5, no. 4, pp. 62-64, 2016.
- [26] M. N. Chaudhari et al., "Face Detection using Viola Jones Algorithm and Neural Networks," *IEEE, Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, pp. 1-6, 2019.
- [27] H. Santoso and R. Pulungan, "A Parallel Architecture for Multiple-Face Detection Technique Using AdaBoost Algorithm and Haar Cascade," *Information Systems International Conference (ISICO)*, pp. 592-597, 2013.
- [28] N. Singhal et al., "Comparative study of machine learning and deep learning algorithm for face recognition," *Jordanian Journal of Computers and Information Technology (JJCIT)*, vol. 07, no. 03, pp. 313-325, 2021.
- [29] Y. A. Ghassabeh et al., "Fast incremental LDA feature extraction," *Pattern Recognition*, vol. 48, no 6, pp. 1999-2012, 2015.
- [30] E. A Khorsheed and Z. A. Nayef, "Face Recognition Algorithms: A Review," *Academic Journal of Nawroz University (AJNU)*, vol. 11, no. 3, pp. 202-207, 2022.

- [31] M. Batta, "Machine Learning Algorithms -A Review," *International Journal of Science and Research (IJSR)*, vol. 9, no. 1, pp. 381-386, 2020.
- [32] R. IHYA et al., "J48 algorithms of machine learning for predicting user's the acceptance of an E-orientation systems," *In Proceedings of the 4th International Conference on Smart City Applications*, pp. 1-8, 2019.
- [33] A. N. Dixit and T. Kasbe, "A Survey on Facial Expression Recognition using Machine Learning Techniques," *IEEE, In proceeding of 2nd International Conference on Data, Engineering and Applications (IDEA)*, Feb. 2020,
- [34] M. Posonia et al., "Machine Learning based Diabetes Prediction using Decision Tree J48," *IEEE, In proceeding of Third International Conference on Intelligent Sustainable Systems*, pp.498-502, 2020.