A Qualitative Performance Comparison Of Supervised Machine Learning Algorithms For Iris Recognition

J. Vasavi¹, M.S. Abirami²

¹Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India.
²Department of Software Engineering, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India.

Abstract: Iris Scanning is the process of biometric authentication that uses classification techniques. Safety is the main issue for humans nowadays. Biometric authentication provides more security because of its uniqueness. Iris recognition is the process of identifying individuals using iris patterns. This study aims to develop a model for iris recognition using Machine Learning methods. The main goal is to progress iris image acquisition, segmentation, classification, Texture Analysis, Feature extraction, cross-sensor recognition, and pattern matching for biometric authentication or verification. This paper reviews a background of iris recognition and literature of recently proposed machine learning methods in different fields of the iris recognition system. The core ideas of various methods and their relationships are investigated to obtain an overview and insights into the development of iris recognition.

Keywords: Machine Learning, Iris Recognition, Acquisition, Classification, Texture Analysis.

1. INTRODUCTION

In recent periods, the utilization of machine vision in the medical investigation is another pattern for huge clinical information entries. In this field, various PC representatives have attempted to apply different procedures to progress the exactness of information arrangement for the given information, grouping methods whose characterization precision will re-establish yield enough data to order the likely development and consequently extemporize the development expectation accuracy. In the present studies, the machine learning algorithms (to name a few, Support Vector Machine, Logistic Regression, Random forest) were applied for iris recognition and obtained with remarkably meaningful results. AutoML revealed that automated techniques can match or improve upon expert human performance in certain machine learning tasks, often in a shorter amount of time [2].

Fig. 1 Eye Image Acquisition

A. Iris
Circular pigmented physiques give color to the eye and also having the capability of controlling the size of the pupil. Iris recognition is the method of biometric authentication of an individual.
B. Pupil
The adaptable opening present at the center of the iris in which the light enters the eye. It is an absorbing pigment in the retina that looks black in color.

C. Sclera
The white outer layer that protects the entire ball of the eye. Its flexibility adds more strength to the eye.

A biometric modal for authenticating a person’s identity using both irises of humans, feature extraction, and matching is proposed to address the problems mentioned. The combination of both face and irises showed greater attention to identification because of the natural connection between the face and irises.

2. STUDY OF SUPERVISED MACHINE LEARNING ALGORITHMS

The vital principle of machine learning is to develop algorithms that can collect input and use statistical analysis to forecast an output [22].

A. Supervised Machine Learning (SML) Algorithms

Machine Learning algorithms are classified as supervised, unsupervised, and semi-supervised algorithms. Here we have a tendency to compare supervised machine learning algorithms to forecast iris recognition which allows us to find the modal for predicting more accurate results.

The following are commonly used Supervised Machine Learning algorithms.

1. Linear Regression
2. Logistic Regression
3. Decision Tree
4. Support Vector Machine
5. Naïve Byes
6. KNN

In the subsequent sections, we briefly label the commonly used supervised ML algorithms.

1) Linear Regression
The relationship between the information factors of two variables x and y is expressed as \( y = a + bx \). Figure 2 shows the plotted x and y esteems for an information set.

![Figure 2 Linear Regression](image)

The Linear regression model was intricate and did not show generally scattered errors, which improved the chances of false predictions. Different regression models differ based on the kind of relationship between dependent and independent variables [4]. Linear Regression is good for learning about the data analysis process with high dimensional dataset [14].

2) Logistic Regression
Logistic regression uses a conversion function \( h(x) = \frac{1}{1 + e^{-x}} \). The probability lies in the range of 0 and 1. The experimental results show that the logistic regression classification model can be used to diagnose dataset quickly, easily, and efficiently, and can help in the medical field [33].
3) Decision Tree

Decision Tree (DT) is a type of supervised learning algorithm mostly used for classification problems. It suits for both categorical and continuous I/O variables. The best classification accuracy was achieved with boosted DT [5]. Decision Trees are found simple to interpret and fast to find out and are a standard element to several indicative protocols[6].

4) Support Vector Machine

Support Vector Machine (SVM) will initially map every knowledge item into associate degree n-dimensions wherever n is that the range of options. SVM showed satisfactory comprehensive performance in all evaluation metrics compared to other SML algorithms [3].

5) Naïve Bayes

The Naïve Bayes model is easy to build and particularly useful for very large data sets. Along with simplicity, Naïve Bayes is known to outperform even highly sophisticated classification methods.

6) KNN

KNN is a non-parametric technique used for classification and regression. The main advantages of the models Naïve Bayes (NB), Decision Tree, Random Forest(RF), and K-Nearest Neighbors are listed below [7]:

1. Misclassified Samples.
2. Outstanding Performance.
3. Uncertainty quantification

3. LITERATURE REVIEW

The algorithms measured in this study are elementary and have been frequently used in other training also.

After acquiring iris images, it undergoes feature vector split up using training and testing dataset, further fed up into several machine learning classifiers such as Support vector machine, KNN, Logistic regression, and random forest for the analysis of classifiers. The result shows that SVM and Random forest provides more accuracy when compared to the other two algorithms [1].

An efficient multimodal biometric authentication system was obtained using parallel architecture when compared with a unimodal system in terms of identification rates using different features. An efficient deep learning approach is used for iris recognition which combines the architecture of Convolutional Neural Network and softmax classifier for extracting discriminatory features of iris images which results in new authentication rates [28].

Among all approaches such as SVM, HACT, hybrid GA, and the XGBoost model, the XGBoost model produced the most accurate prediction performance [8]. classification in specific is similar to SVM, decision trees, and Bayesian networks [9], but SVM provides the best classification result.

ML-based approaches have the potential in the classification and identification of Non-Functional Requirements [10]. Among the seven SML, the results show that ANN has the best prediction accuracy[11].

Supervised machine learning is also emerging to optimize prediction accuracy [12]. The comparative results say that the K-neighbor regression model is the best model for prediction with the given data set [13]. This training affords a summary of the virtual presentation of dissimilar alternates of SML for disease prediction [15].

ML can enhance the diagnostic and prognostic capacity of more traditional regression techniques. when analyzing the data in RapidMiner and applying the same model, the best algorithm was SVM[16].

Random Forest (RF), extreme gradient boosting XGBoost SVM, and Multilayer Perceptron (MLP). The author says that the best result was produced by XGBoost [17].SVM outperforms other classifiers, in term of accuracy and duration spent on training process[18].

In terms of image processing, the use of deep learning for enlightening overall image excellence is at the forefront of the current developments [19]. The classification approach applying the CNN-based
deep learning model can then expand the efficacy and proficiency of complaint handling on the given data set [20].

SML requires a large number of labeled data, semi-supervised learning has been established to leverage unlabeled data as well as labeled data for classification [21]. With the use of homogeneous data sets over unstandardized data sets, the performance of the training algorithm is improved, thereby reducing complication and calculation time [23].

This paper provides a successful application of unsupervised and supervised machine learning in the data set, perhaps help develop more effective combination strategies [24].

Within ML approaches, supervised methods outperformed unsupervised. The best classification method in terms of quality and completeness was SVM [25]. The author shows in his work that the random forest produces better results among other cited ML algorithms [26].

The roles of this paper are all about SML and do not include unsupervised learning and Reinforcement Learning. Moreover, ML techniques become more accurate with bigger datasets [27].

SVM algorithm used a high dimension to classify the observation and also by adding extra pre-processing the accuracy rate can be enhanced [29].

We can conclude that ANN is performing better in forecasting the exchange rate when compared to multilinear regression [30].

The most popular algorithm J48 is used to categorize dissimilar applications and perform accurate results of the classification to examine the data categorically and unceasingly [31].

In recent years, with the rapid development of ML, SML plays an important part that has received extensive attention in the ground CS for estimation [32].

The results of this article encourage decision architects to test and consider the use of ML models in their databases [34]. SML is quite common in classification problems, therefore, ML algorithms require accuracy, precision, and least error to have SML for prediction [35]. The calculation of the mean accuracy of experimental networks has shown better results using ANN [36].

Naïve Bayes and Random Forest classification algorithms were found to be the next accurate after SVM accordingly [37]. SML is capable of solving linear and nonlinear classification problems [38].

This article provides that all algorithms acquire some kind of outlines from the training dataset and apply them to the test dataset for prediction or classification [40].

DT and NB classification algorithms are used to perform dataset classification most accurately without any loss of information [41].

To analyze the experiments, SVM minimizes the generalization error which is obtained by increasing the marginal distance between the data to obtain the decision [39].

Among all biometric traits such as Face biometric, Fingerprint biometric, Retina biometric, signature biometric, iris biometric provides more attention for recognizing humans. [43,44].

The experimental results showed that segmentation and localization schemes detect and reduce the noises around the iris portion of the eye. [42]

The model reached perfect classification accuracy, eliminates the false acceptance rates using machine learning methods by converting the normalized image into one dimensional set that undergoes stylometric feature extraction [45].

CNN features are used to find the generic objects that represent iris images which effectively extracting features and achieving promising results. [46]

The face recognition system uses Kinect hardware devices that perform human face detection and localization, which extracting features by local binary patterns. [47].

Sclera images are referred to get the grouped features and fed up into the Support Vector machine to train an identity classifier [48].

Some of the normalized features are extracted from iris images and then, they are classified using Support Vector Machine and Neural Networks. The result showed that the Support Vector Machine provides more recognition rate when compared to the backpropagation neural network [49].
Multilayer Perceptron Neural Network and Particle Swarm Optimization algorithms are used for the classification of human iris images. The data are normalized, trained, and tested using these algorithms and produced better results for iris recognition [50].

The features extracted from the periocular region and iris are applied for iris recognition using a Multimodal system approach which helps in the creation of a truthful system and provides better results for recognition [51].

4. CLASSIFICATION ARCHITECTURE
A. Eye Image Acquisition
The eye image acquisition is the data that we use to model the algorithm for getting the desired output. This data will be enriched with data footnote. The performance of machine learning methods may increase if you feed enough data. The usage of a large number of input data will provide more accuracy.

B. Normalization
Normalization is nothing but a deformation of iris images such as translation, angular scaling, overall scaling, etc to make the iris texture to be aligned. Thus, each point in the given image is remapped for feature extraction. Segmentation provides a way for finding the exact outline of the object present in an image. Supervised machine learning algorithms initially hired the dataset to coach the algorithmic program. This trained algorithmic program is then consumed the data to compare them into similar data.

C. Feature Extraction
When we use a similar dataset for both training input and testing the data, it helps us to minimize the data inconsistencies and provides the better characteristics of the method to understand easily. Once the method has been modeled by using the training dataset, we test the model by making predictions.
against the test data. Feature extraction is used to reduce the initial input data into meaningful groups for processing.

D. Classification

The classification is used to find the model that explains and differentiates data classes and ideas. Classification is the concept of identifying a different category based on input data. It allows us to get new observations based on training input data and testing data. Also, we need to check the feasibility of the work. In such cases, a classifier is a must for predicting class labels and testing the dataset using the model predicted to provide accurate results. Classification provides a way to find “what is present inside the image?” The images can be classified into different categories.

E. Recognition Rate

Classification provides a way of identifying an authorized person. Biometric authorization gives access to biometric traits. The biometric system allows the person who is authorized to access the system data, otherwise it rejects the unauthorized person.

5. EXISTING EVALUATION METHODS

<table>
<thead>
<tr>
<th>REFERENCE NO.</th>
<th>YEAR</th>
<th>JOURNAL</th>
<th>METHODOLOGY</th>
<th>DATASET</th>
</tr>
</thead>
<tbody>
<tr>
<td>[1]</td>
<td>2019</td>
<td>IEEE</td>
<td>Machine Learning algorithm</td>
<td>MMU-1 Iris Dataset</td>
</tr>
<tr>
<td>[3]</td>
<td>2020</td>
<td>Elsevier</td>
<td>Supervised Machine Learning Algorithm</td>
<td>Nine healthy participants were recruited</td>
</tr>
<tr>
<td>[8]</td>
<td>2019</td>
<td>Elsevier</td>
<td>ML Algorithms</td>
<td>64 companies</td>
</tr>
<tr>
<td>[9]</td>
<td>2019</td>
<td>Indian Journal of Applied Research</td>
<td>SVM, DT, and BN</td>
<td>Cancer, Brain Tumour, ENT Ailment, Heart Disease, Thyroid</td>
</tr>
<tr>
<td>[12]</td>
<td>2019</td>
<td>Elsevier</td>
<td>Supervised learning, Unsupervised learning, Reinforcement learning</td>
<td>Mental Disorders</td>
</tr>
</tbody>
</table>
### 6. CONCLUSION

We conclude from the analysis of recent research that Supervised machine learning algorithms provide a concise model for the distribution of class labels in terms of predictor features. Also, in specific, the four main supervised machine algorithms namely Support vector machine, Linear Regression, Logistic Regression, and decision tree are used to obtain a methodology to analyze iris image acquisition, segmentation, and Texture analysis and to realize their automatic classification and clustering of data. It also provides a way to find the modal for iris pattern recognition.
7. REFERENCES


