Effective Classifier Identification in Biometric Pattern Recognition

S. M. Emdad Hossain
Department of Information Systems, College of EMIS, University of Nizwa, Oman
emdad.hossain@unizwa.edu.om

Sallam O. F. Khairy
Department of Information Systems, College of EMIS, University of Nizwa, Oman
sallam@unizwa.edu.om (corresponding author)

Arockiasamy Soosaimanickam
Department of Information Systems, College of EMIS, University of Nizwa, Oman
arockiasamy@unizwa.edu.om

A. M. Raisuddin
Military Technological College, Oman
muhammad.ahmed@mtc.edu.om

Received: 5 April 2024 | Revised: 22 July 2024 | Accepted: 26 July 2024

 DOI: https://doi.org/10.48084/etasr.7424

ABSTRACT

Next-generation identity verification using biometric features is nearly foolproof with the right classifier. However, selecting the correct classifier poses a key challenge, particularly in the recognition of biometric patterns. High-potential projects may face delays due to a lack of the right recognition mechanism or the malfunction of the selected classifier. This could also result from not choosing the appropriate classifier that aligns with the project's patterns. This study aims to evaluate various classifiers with potential in biometric research and the capabilities of different machine learning algorithms. Several classifiers were experimentally evaluated in combination with dynamic algorithms. The ultimate objective was to identify a standard classifier suitable for general biometric pattern recognition. Using well-known biometric pattern datasets, multivariate algorithms, such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA), were applied. These methods were combined with different classifiers, including SVM-L, MLP, KNN, etc. After analyzing the results obtained, the combination of LDA with MLP outperformed other approaches in terms of accuracy.

Keywords-classifier; challenge; pattern; algorithm; biometric

I. INTRODUCTION

Biometric pattern recognition is a very popular and proven approach to identifying human identity. Today, researchers are concerned with robustness in the areas of methods, data selection, classifier selection, etc. Several different algorithms are available, but they are not equally suitable for applications in advanced identification systems. There are many arguments, such as effectiveness, applicability, algorithm or classifier structure, nature and objective of the system, etc. Eventually, with all these arguments, it is well established that classifiers, algorithms, and data quality are still essential. This is due to various reasons, as the success parameters or standards of a system differ from others. For instance, there was a time when researchers believed that around 90% and above would be good enough to rate a system as successful.

Taking into account all potential aspects, this study performed a set of experiments to provide a clear direction in selecting a biometric pattern classifier to ensure robust identification to meet the needs of the national and international biometric security research communities. Several well-known classifiers, i.e., Sequential Minimal Optimization (SMO), Support Vector Machine (SVM), SVM-Poly, SVM-RVF, SVM-Sigmoid, Bayesian Linear, Bayesian Quadratic, Multilayer Perceptron (MLP), KNN, J48, Naïve Bayes, etc. were examined and evaluated. The evaluation results clearly indicate that MLP is one of the most effective classifiers that were examined.

II. BACKGROUND WITH APPLIED CLASSIFIERS

Biometric pattern recognition is one of the key areas that the research community investigates to enhance robustness, due to the increasing demand for robust biometric security systems. Many studies have highlighted the need for biometric security systems. Continuous research is carried out on different aspects of biometric-based security systems [2]. As researchers are moving forward, certain limitations and obstacles are coming up. In some cases, the classifier choice is one issue that poses challenges. This study focuses on biometric feature classifiers, investigating several classifiers for this purpose, along with their applicability, formation, usability, and effectiveness.

A. SVM and SMO

SVM and SMO operate similarly, as they share common principles in classification. SVM classifiers perform classification through a multidimensional approach that separates each data point into distinct class categories. This process typically involves the creation of hyperplanes during experimentation. An SVM constructs one or more hyperplanes in a high-dimensional or infinite-dimensional space, which can be applied for tasks such as subgrouping, reducing dimensionality, or similar operations. Effective separation is achieved by the hyperplane with the greatest distance to the nearest training data point of any class, referred to as the functional margin. This is due to the commonly applied principle that a greater margin size leads to reduced generalization errors for a specific classifier. However, in cases where the sets to be distinguished are not linearly separable in the original finite-dimensional space, SVM maps this space into a substantially higher-dimensional one, aiming to simplify separation. To maintain reasonable computational efficiency, SVM techniques intentionally design mappings that facilitate the computation of dot products in terms of the original space variables. This is achieved by defining these mappings in a kernel function chosen to be suitable for the problem at hand [3].

On the other hand, Sequential Minimal Optimization (SMO) is a Support Vector Machine (SVM) classifier that utilizes sequential minimal optimization for learning. SMO breaks down the overall Quadratic Programming (QP) problem into smaller subproblems, ensuring convergence using Osuna's theorem [4]. Unlike other methods, SMO solves the smallest possible optimization problem at each step. The advantage of SMO lies in its ability to obtain analytical solutions for multiple-instance multipliers. Additionally, SMO does not require additional matrix storage. SMO comprises two main components: an analytical technique to determine the two Lagrange multipliers and an empirical method to select the optimal multipliers to achieve the best optimization [4].

$$y1 \;/\!=\; y2 \;=>\alpha1-\;\alpha2\;=\;k \qquad (1)$$

$$y1\;=\;y2\;=>\;\alpha1+\alpha2\;=\;k \qquad (2)$$

Nevertheless, the multi-instance multipliers must satisfy all the constraints of the complete problem. Due to the linear equality constraint, these multipliers are restricted to lie on a diagonal line. Consequently, during an iteration of SMO, it becomes necessary to locate an optimal solution for the objective function along a diagonal line segment [4].

B. Bayesian Linear and Quadratic

Both classifiers are quite old but still very effective in the biometric classification arena. Bayesian linear and quadratic discriminant classifiers use Bayesian decision rules to classify learned feature vectors. The linear classifier applies a unified covariance estimate when fitting multivariate normal densities for each group, whereas the quadratic discriminant classifier employs separate covariance estimates. Both methods rely on likelihood ratios to assign observations to their respective groups. These classifiers are effective tools for the classification of learned feature vectors, distinguished by covariance estimation, as the linear classifier assumes shared covariance and the quadratic classifier allows for distinct covariance estimates per group, offering flexibility in modeling complex data distributions [5]. The Bayesian decision rule states that when faced with a set of classes M, characterized by known parameters in model Ω, a collection of extracted feature vectors X is assigned to the class with the highest probability. This principle, encapsulated in (1), is widely known as the Bayesian decision rule.

$$P(\omega_i|\bar{x}) = \frac{p(\bar{x}|\omega_i)\times P(\omega_i)}{p(\bar{x})} \qquad (3)$$

The a-posterior probability is computed using the Bayesian law of statistics. Assuming that the features follow a normal distribution leads to the formulation of a quadratic classifier known as the Bayesian quadratic classifier. The model Ω includes the mean and covariance of the training vectors, and the likelihoods are evaluated based on the aforementioned approach. Furthermore, KNN [6] also works like Bayesian and looks for the nearest neighbor data point.

C. J48 Classifier

Also known as the C4.5 algorithm [7], this is a decision tree generation technique developed by Ross Quinlan, which is an evolution of his earlier ID3 algorithm. C4.5 is widely used to create decision trees, primarily for classification purposes, and is recognized as a statistical classifier. Like ID3, C4.5 builds decision trees from a training dataset utilizing the concept of information entropy. The J48 classifier exhibits high classification accuracy, particularly when combined with superior algorithms. The training data, denoted as a set $S = s_1, s_2$, consist of pre-classified samples, where each sample $s_i = x_1, x_2$ represents a vector of attributes or features. Additionally, these data include a vector $C = c_1, c_2$, indicating the class membership of each sample [8]. At each node in the tree, the C4.5 algorithm selects the attribute that best divides the sample set into subsets, with each subset exhibiting a stronger presence of one class over others. The criterion for attribute selection is the normalized information gain, which quantifies the change in entropy resulting from using an attribute for data partitioning. The attribute with the highest normalized information gain is chosen to make the decision. The C4.5 algorithm is then applied recursively to the resulting smaller sublists [9].

D. Other Classifiers

MLP is a well-known classifier in the same line of neural networks. An unsophisticated Bayes classifier is a simple probabilistic classifier that employs Bayes' theorem while making the fundamental assumption of independence (often referred to as uninformed assumptions). A more complex representation of the underlying probability model is known as an independent feature model. In simpler terms, an uninformed Bayes classifier assumes that the presence or absence of a particular feature in a class is unrelated to the presence or absence of any other feature, given the class variable. In essence, it operates in a straightforward and easily understandable manner [8].

$$Posterior = \frac{Prior \times Likelihood}{Evidence} \quad (4)$$

All model parameters, including class priors and feature probability distributions, can be approximated using the relative frequencies observed in the training dataset, representing maximum likelihood probabilities. The class prior can be calculated by assuming equal probabilities for all classes, i.e., $priors = 1 / (number\ of\ classes$, or by calculating the probability estimate for each class based on the number of samples in that class divided by the total number of samples. Estimating parameters for feature distributions requires assuming a specific distribution or generating nonparametric models based on observed features in the training set [8].

III. METHODOLOGY

The methods were carefully selected, considering the importance and sensitivity of the field. A variant of Principle Component Analysis (PCA), known as V-PCA, was employed [11]. V-PCA constitutes a fundamental technique in feature learning that enables the automated discovery of concise and meaningful data representations, without relying on intricate feature extraction methods or domain-specific expertise. It stands as a well-established approach for data de-correlation and dimensionality reduction. In particular, it effectively concentrates the primary variance of the initial data within a lower-dimensional subspace defined by eigenvectors and eigenvalues. By projecting the original data onto this subspace, the feature representation is obtained, and the optimal target dimensionality of the feature space is determined through automated analysis of the eigenvalue spectrum of the sample covariance.

However, when input data lack proper normalization, the performance of PCA features [12] often fails to meet expectations. Blind-range normalization proves ineffective in addressing this issue, particularly when the components relate to various aspects of a phenomenon. This challenge is particularly pronounced in the context of gait recognition from multiple viewpoints. To overcome this obstacle, an alternative representation, rooted in the empirical cumulative distribution function is devised: ECDF [13] of the gait silhouette/contour (x, y) for each frame. This representation remains independent of absolute ranges while preserving essential structural information. In this process, the silhouettes for each frame within the gait image sequence are extracted using background subtraction, with the boundary contours serving as the gait silhouette.

On the other hand, Deep Learning Feature (DLF) has emerged as a powerful tool in the field of biometric pattern recognition [14]. It stands as a robust method for semi-supervised feature discovery, leveraging autoencoder networks designed to learn a lower-dimensional representation of input data while minimizing reconstruction errors. This approach incorporates feedforward neural networks characterized by a single input layer, an output layer, and an odd number of hidden layers. Each layer maintains full connectivity with its adjacent layers and is governed by a nonlinear activation function. During the training process, the primary objective is to reconstruct the input data in the output layer. As data traverses the network layers during encoding, the autoencoder gains proficiency in capturing nonlinear feature representations.

To ensure robust training of the model, according to [15], the layers of the autoencoder network are systematically acquired using a bottom-up approach known as a greedy strategy. In this method, each consecutive pair of layers within the encoder is treated as if it were a Restricted Boltzmann Machine (RBM). An RBM, a two-layer graphical model characterized by full connectivity and a bipartite structure, can generatively model data. It trains a set of hidden stochastic binary units, which serve as detectors for low-level features. During the training of each successive layer pair, the activation probabilities of the feature detectors from one RBM are employed as input data for the subsequent RBM in the sequence.

In RBMs, different techniques are applied to represent real-valued input units. In the initial RBM layer, Gaussian visible units are employed to activate binary stochastic feature detectors, creating a Gaussian-binary configuration. As moved to subsequent layers, the standard binary-binary RBM configuration is used, and the final layer utilizes a binary linear RBM, which essentially carries out a linear projection. The training procedure involves processing data samples in batches, with the aim of ideally including samples from all classes that are part of the training dataset. In particular, RBMs are trained in an unsupervised manner, meaning that class-related information is not a mandatory part of the training process.

To assess the efficacy of feature learning methods, including V-PCA and DLF features, for gait-based human identity recognition, a series of experiments were conducted using various data subsets sourced from the UCMG database [16], which is one of the largest biometric security research databases in the world. For baseline comparisons, standard PCA and LDA features were also extracted. The performance of these features, combined with the proposed classifiers, underwent a comprehensive evaluation.

IV. EXPERIMENTS AND ANALYSIS

A total of three sets of experiments were conducted using various subsets of the UCMG dataset [16]. Tables I, II, and III and Figures 1, 2, and 3, present the performance results for each set of experiments. These results are measured in terms of recognition accuracy and various statistically significant performance metrics, including True Positive Rate (TPR), False Positive Rate (FPR), recall, precision, and F-score.

TABLE I. PERFORMANCE OF MLP CLASSIFIER

Algorithm/Classifier	Correct identification (%)
vPCA-MLP	51.57
DLF-MLP	78.4
LDA-MLP	100

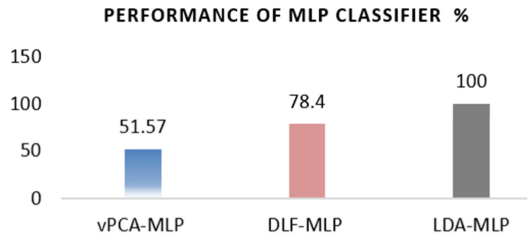

Fig. 1. Performance of the MLP classifier.

Using feature learning techniques, namely LDA, V-PCA, and DLF, the gait feature vectors were extracted from silhouette images that portrayed individuals walking in each video sequence. Identification experiments were carried out in two modes: a single mode and a multimodal fusion mode. To assess feature performance in less cooperative camera conditions, a variety of views were used for both training and testing. To achieve this, a fused training template was created by combining features from different views, while the test data were sourced from a view that was not utilized in constructing the fused training template. This approach was imperative to avoid a significant increase in error when training data from one view and test data from another. For further investigation on the use of biometric pattern recognition, studies in [17-19] can be used.

In all experiments, a reduced dataset encompassing 10 classes (representing 10 individuals) and 20 features (comprising PCA, V-PCA, LDA, and DLF) were employed. Initial experimentation revealed that approximately 95% of variations could be effectively captured with around 10 features, and expanding the dimensionality did not yield substantial performance enhancements. Both LDA and DLF exhibited outstanding performance. Furthermore, a subsequent set of experiments was carried out to assess whether alternative established classifiers employing different kernels would yield improved results, as demonstrated in Table III. Remarkably, for both datasets, the straightforward NB classifier surpassed more complex SVM classifiers across various kernel types, possibly due to to the superior learning capability of DLF features.

TABLE II. PERFORMANCE OF N-BAYES AND SVM-L

Classifier	Database	Accuracy (%)	TP	FP	Precision	Recall	F-score
NB	UCMG	92.25	0.92	0	0.93	0.92	0.92
SVM-L	UCMG	78.75	0.78	0	0.81	0.79	0.78

TABLE III. PREFERENCE OF FUSION IN LDA AND DLF

Classifier	Database	Accuracy (%)
NN	UCMG	78.75
MLP	UCMG	100%
SVM-L	UCMG	27.63

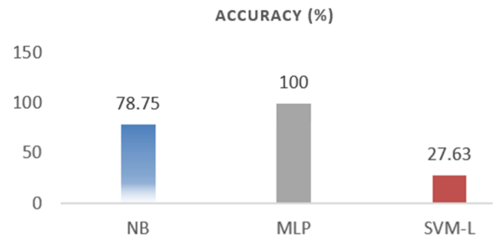

Fig. 2. Preference of fusion in LDA and DLF.

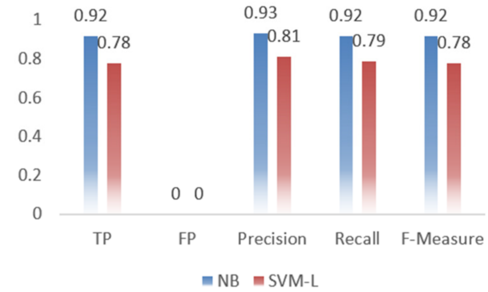

Fig. 3. Performance of Naïve Bayes and SVM-L classifiers.

Fusion with LDA and DLF showed a significant improvement. As shown in Table III, NN achieved relatively poor results compared to MLP, which achieved 100% detection accuracy. In addition, SVM-L showed very poor accuracy compared to NB and MLP. After a close view of all experiments using the stated algorithms and classifiers, it seems that the dataset from the UCMG database is working extremely well in detecting or classifying people from a distance using low-resolution videos or images. Likewise, it is also clear that the combination of LDA-MLP achieved the maximum detection accuracy rate.

V. CONCLUSION

This study used V-PCA, LDA, and DLF for dimensionality reduction and feature extraction. A variety of different classifiers was applied to identify a person from the extracted features. MLP consistently showed high effectiveness and provided excellent accuracy for correct identification/detection, including both true and false positives. Furthermore, to compare accuracy and robustness, a full set of experiments was carried out in different dimensions with various well-known classifiers. All experiments were carried out using identified data from the UCMG database. Initially, the results were relatively poor, compared to the final results, due to the feature extraction algorithm. The results show that LDA and MLP is the best combination for identifying a person from low-resolution surveillance video. Future work will involve further experiments for practical implementation in public spaces, even with low-resolution surveillance videos or images.

REFERENCES

[1] C. L. Wilson, *Biometric Accuracy Standards.* National Institute of Science and Technology (NIST), 2003.

[2] J. Fierrez, A. Morales, R. Vera-Rodriguez, and D. Camacho, "Multiple classifiers in biometrics. part 1: Fundamentals and review," *Information Fusion*, vol. 44, pp. 57–64, Nov. 2018, https://doi.org/10.1016/j.inffus.2017.12.003.

[3] C. M. Travieso, J. B. Alonso, and M. A. Ferrer, "Facial identification using transformed domain by SVM," in *38th Annual 2004 International Carnahan Conference on Security Technology, 2004.*, Albuquerque, NM, USA, Oct. 2004, pp. 321–324, https://doi.org/10.1109/CCST.2004.1405412.

[4] J. Platt, "Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines," Microsoft, MSR-TR-98-14, Apr. 1998.

[5] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*. Academic Press, 2006.

[6] O. Harrison, "Machine learning basics with the K-nearest neighbors algorithm," *Towards data science*, vol. 11, Sep. 2018.

[7] N. Saravanan and V. Gayathri, "Performance and classification evaluation of J48 algorithm and Kendall's based J48 algorithm (KNJ48)," *International Journal of Computer Trends and Technology*, vol. 59, no. 2, pp. 73–80, 2018.

[8] J. R. Quinlan, *C4.5: Programs for Machine Learning*. Morgan Kaufmann, 2014.

[9] A. Hussain, G. Ali, F. Akhtar, Z. H. Khand, and A. Ali, "Design and Analysis of News Category Predictor," *Engineering, Technology & Applied Science Research*, vol. 10, no. 5, pp. 6380–6385, Oct. 2020, https://doi.org/10.48084/etasr.3825.

[10] Y. Chen and J. C. Fondeur, "Biometric Algorithms," in *Encyclopedia of Biometrics*, S. Z. Li and A. Jain, Eds. Boston, MA, USA: Springer US, 2009, pp. 64–68.

[11] S. Ruggieri, "Efficient C4.5 [classification algorithm]," *IEEE Transactions on Knowledge and Data Engineering*, vol. 14, no. 2, pp. 438–444, Mar. 2002, https://doi.org/10.1109/69.991727.

[12] L. I. Smith, A tutorial on Principal Components Analysis, Feb. 2002.

[13] B. Sukresno, A. Hartoko, B. Sulistyo, and Subiyanto, "Empirical Cumulative Distribution Function (ECDF) Analysis of *Thunnus.sp* Using ARGO Float Sub-surface Multilayer Temperature Data in Indian Ocean South of Java," *Procedia Environmental Sciences*, vol. 23, pp. 358–367, Jan. 2015, https://doi.org/10.1016/j.proenv.2015.01.052.

[14] K. Pearson, "LIII. On lines and planes of closest fit to systems of points in space," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, Nov. 1901, https://doi.org/10.1080/14786440109462720.

[15] G. E. Hinton, "To recognize shapes, first learn to generate images," in *Progress in Brain Research*, vol. 165, pp. 535–547, 2007.

[16] G. Chetty, "Multimodal Biometric Gait Database: A Comparison Study," *Journal of Next Generation Information Technology*, vol. 5, no. 4, pp. 71–82, 2014.

[17] Y. Said, M. Barr, and H. E. Ahmed, "Design of a Face Recognition System based on Convolutional Neural Network (CNN)," *Engineering, Technology & Applied Science Research*, vol. 10, no. 3, pp. 5608–5612, Jun. 2020, https://doi.org/10.48084/etasr.3490.

[18] S. M. Hassan, A. Alghamdi, A. Hafeez, M. Hamdi, I. Hussain, and M. Alrizq, "An Effective Combination of Textures and Wavelet Features for Facial Expression Recognition," *Engineering, Technology & Applied Science Research*, vol. 11, no. 3, pp. 7172–7176, Jun. 2021, https://doi.org/10.48084/etasr.4080.

[19] A. O. Aljahdali, F. Thabit, H. Aldissi, and W. Nagro, "Dynamic Keystroke Technique for a Secure Authentication System based on Deep Belief Nets," *Engineering, Technology & Applied Science Research*, vol. 13, no. 3, pp. 10906–10915, Jun. 2023, https://doi.org/10.48084/etasr.5841.