IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received October 8, 2020, accepted October 13, 2020, date of publication October 30, 2020, date of current version January 8, 2021.

Digital Object Identifier 10.1109/ACCESS.2020.3035110

A Hybrid Model Combining Learning Distance Metric and DAG Support Vector Machine for Multimodal Biometric Recognition

IBRAHIM OMARA[1,2], AHMED HAGAG[3], (Member, IEEE), SOULEYMAN CHAIB[4], GUANGZHI MA[1], FATHI E. ABD EL-SAMIE[5,6], (Member, IEEE), AND ENMIN SONG[1], (Senior Member, IEEE)

[1]School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China
[2]Department of Mathematics and Computer Science, Faculty of Science, Menoufia University, Menoufia 32511, Egypt
[3]Department of Scientific Computing, Faculty of Computers and Artificial Intelligence, Benha University, Benha 13518, Egypt
[4]LabRI-SBA Lab, Ecole Superieure en Informatique, Sidi Bel Abbès 22016, Algeria
[5]Department of Electronics and Electrical Communications, Faculty of Electronic Engineering, Menoufia University, Menoufia 32952, Egypt
[6]Department of Information Technology, College of Computer and Information Sciences, Princess NourahBint Abdulrahman University, Riyadh 21974, Saudi Arabia

Corresponding author: Guangzhi Ma (maguangzhi@hust.edu.cn)

This work was supported in part by the Higher Education Commission of Egypt and the Chinese Government, in part by the National Key Research and Development Program of China under Grant 2017YFC0112804, and in part by the National Natural Science Foundation of China under Grant 81671768.

ABSTRACT Metric learning has significantly improved machine learning applications such as face re-identification and image classification using K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) classifiers. However, to the best of our knowledge, it has not been investigated yet, especially for the multimodal biometric recognition problem in immigration, forensic and surveillance applications with uncontrolled ear datasets. Therefore, it is interesting and very attractive to propose a novel framework for multimodal biometric recognition based on Learning Distance Metric (LDM) via kernel SVM. This paper considers metric learning for SVM by investigating a hybrid Learning Distance Metric and Directed Acyclic Graph SVM (LDM-DAGSVM) model for multimodal biometric recognition, where LDM and DAGSVM are two emerging techniques in dealing with classification problems. Different from existing multimodal biometric recognition methods, the proposed approach aims to learn Mahalanobis distance metric via kernel SVM to maximize the inter-class variations and minimize the intra-class variations, simultaneously. Experimental results on the uncontrolled datasets such as AR face and AWE ear datasets show that the proposed approach achieves competitive performance compared with models working on individual modalities and overperforms the state-of-the-art multimodal methods. The proposed model achieves five-fold classification accuracy around 99.85 % for the face and ear images.

INDEX TERMS Biometrics, multimodal biometrics, face and ear images, Mahalanobis distance, metric learning, DAGSVM.

I. INTRODUCTION

Unimodal biometric authentication systems got more attention in the last decades for intelligent applications such as Internet of Things (IoTs), Automated Teller Machine (ATM), surveillance, immigration, and mobile applications. However, some unimodal biometric authentication systems are potentially vulnerable to forgery, which means that the biometric system can be cheated [1]. For example, fingerprint is the most popular biometric trait due to its perceived uniqueness and persistence [2]. However, the Apple iPhone Touch ID fingerprint reader might be cheated by using non-authentication fingers [3]. Furthermore, the big challenging issue for unimodal biometric recognition systems is to select discriminative features to accomplish personal authentication in the presence of large variations among biometric samples

The associate editor coordinating the review of this manuscript and approving it for publication was Andrea F. Abate.

for the same person. Thus, one type of features is usually neither efficient nor sufficient to predict the right subject, especially for the collected images in various conditions such as illumination, rotation, and occlusion conditions. Therefore, most researchers pay more attention to multimodal biometric recognition to increase identification performance and to provide more security. To deal with multimodal biometric recognition, feature fusion has become a very important research aspect [4]–[12], because fusing different types of features provides complementary information. Most of the recent human recognition works [10], [13–[17] utilized feature level fusion to overcome the challenges of constrained resources, and to increase the system security and system performance. The well-known traditional feature fusion methods involve serial and parallel fusions. Serial fusion can apply by adding or concatenating two or more feature vectors to a single feature vector; on the other hand, parallel fusion is more difficult and hard in utilizing more than two features, due parallel fusion provides a complex vector by collecting a feature set as the imaginary component and another feature sets as the real component, which makes its application limited.

Recently, recognition metrics have been considered as an important research point for the development of multimodal biometric recognition systems. To the best of our knowledge, most of the previous works [4], [9], [10], [13], [15], [16] have adopted traditional distance metrics and classifiers. Due to the number of images per person that are usually limited to 3∼5 images and with noises, traditional distance metrics and classifiers cannot achieve the performance desired. Moreover, they are sensitive to noises. To deal with classification problems and achieve better generalization ability, metric learning for SVM is considered. As the LDM and the DAGSVM are two emerging techniques in dealing with classification problems and can achieve better generalization ability than the traditional distance metrics and classifiers, the LDM-DAGSVM model for the multimodal biometric recognition system is investigated.

In addition, there is no reported work based on LDM for SVM. Learning models have got much more attentions in the last decades especially for machine learning applications such as classification applications [18], [19], [20]–[23], face recognition [24], [25], and human re-identification [26]. Metric learning aims to learn a valid distance from the given training data or a similarity function for a given problem. However, most existing metric learning methods are based on convex and non-convex optimization algorithms or multiple kernel classification. In this paper, we investigate a kernel classification model that can be used to improve the state-of-the-art multimodal biometric recognition system based on metric learning algorithms. In addition, multimodal biometric recognition systems can be enhanced using metric learning approaches, which utilize various known efficient classifiers such as K-Nearest Neighbor (K-NN) and SVM. Therefore, the main motivation of our proposed multimodal biometric recognition system includes two aspects: face and ear images representation, and classification. This paper exploits the advantages of local features fusion to represent face and ear images, by using Discriminant Correlation Analysis (DCA) in feature fusion algorithm that enhances the efficiency and effectivity [11], [15]. The proposed work presents kernel DAGSVM to improve the SVM performance based on learning the Mahalanobis distance using Radial Basis Function (RBF) for a multimodal biometric recognition system.

Hence, a multimodal biometric recognition system based on LDM is investigated in this paper to achieve higher classification rates for face and ear images. The experiments are performed on biometric applications of Annotated Web Ear (AWE) dataset, Mathematical Analysis of Images (AMI) ear dataset, and the Georgia Tech, Olivetti Research Laboratory (ORL) face datasets, and AR face dataset. The experimental results indicate that multimodal of individual modalities can improve the overall performance of the human biometric recognition system, even in the case of low-quality data. The results also demonstrate that the proposed model performs better than classical and traditional multimodal biometric models. The adopted LDM-DAGSVM model is particularly useful for two reasons. First, it achieves comparable classification accuracies with those of the state-of-the-art methods, and it clearly outperforms the other multimodal biometric methods not explicitly geared towards LDM. In addition, it provides the researchers in the multimodal biometric recognition systems a convenient way for using various metric learning algorithms via SVM. The major contributions of this work are five-fold.

1. This paper presents a novel framework for multi-biometric recognition through LDM and kernel SVM. Few previous works have studied biometrics in the context of metric learning, and to the best of our knowledge, no prior works have attempted LDM for multimodal biometric recognition.
2. To optimize the authentication process, we combine LDM with DAGSVM, which are two emerging techniques in dealing with classification problems. This approach aims to maximize the inter-class variations and minimize the intra-class variations between biometrics.
3. Furthermore, to improve KNN and SVM performance, this work learn the Mahalanobis distance via DAGSVM. It outperforms the previous works based on SVM or KNN, such as learning Mahalanobis distance through the KNN.
4. Discriminative and commonly-used standard features such as Local Binary Patterns (LBPs) and Histogram of Oriented Gradients (HOGs) for face and ear images are used to represent each trait. Besides, DCA algorithm is exploited to combine and reduce features vectors dimension for face and ear images, separately.

5 We illustrate that multimodal biometric recognition can be robust and efficiently implemented based on LDM through a kernel classifier such as the SVM. Experimental results show that the proposed multimodal

biometric recognition model outperforms other state-of-the-art methods.

In addition, this work indicates a new direction for multimodal biometric recognition problem by developing new systems using metric learning approaches with various known active classifiers such as KNN and SVM. This paper is organized as follows. In Section II, we introduce a brief review of the related work. In Section III, we present the proposed model in detail. In Section IV, we evaluate and compare the proposed model with other state-of-the-art methods on three complicated datasets. Finally, the conclusion and future works are presented in Section V.

II. RELATED WORK

Automated biometrics authentication refers to the automated human recognition based on physiological or/and behavioral characteristics. Biometric authentication task is known as predicting whether two biometric traits belong to the same person or not. Therefore, human biometrics authentication has received more increasing attention recently in many intelligent applications such as IoT, Mobile application, ATM, and surveillance. Human biometric traits include physiological traits like face, fingerprint, palm print, iris, and ear, or/and behavioral traits such as gait, keystrokes, voice, and signature. Face and ear recognitions have received a lot of attention in the last decade as these two traits are proved to be the promising biometric traits due to their uniqueness, collectible, permanence and universality [2], [27]. However, they have unique advantages and also have some limitations [28], [29]. Face and ear images have many advantages such as both face and ear traits are large and visible for acquisition, that means they are passive, non-intrusive traits, and can be collected within one sensor. However, face and ear recognition systems have many challenges such as facial expression, makeup, mask, rotation and occlusion from hair, glasses, or rings. On the other side, the human ear images have complementary information for face image. The human ear has stable structure through expression and age, being visible and large, that means it can be collected without user cooperation; furthermore, human ears for identical twins and triplets are different [30].

Chang *et al.* [28] presented a multimodal biometric recognition system for human identification based on standard Eigen-faces and Eigen-ears to represent face and ear images, respectively. They adopted feature-level fusion based on a simple concatenation that may trigger the dimension problem. The authors in [9] have motivated the utilization of Kernel Principal Component Analysis (KPCA) to solve the dimension problem with a scanty accuracy around 94.5% [28]. Therefore, several attempts [10], [15], [31]–[34] have presented multimodal biometric recognition systems based on face and ear images with various level fusion to improve the recognition accuracy and the system performance. To overcome the above-mentioned limitations, the authors of [34] and [31] have exploited 3D ear images and sparse representation for classification, respectively. Mahoor *et al.* [34] have fused 3D ear and 2D face images to build a multimodal biometric recognition system. Active shape model and Gabor filter were used for extracting landmarks from face images. Then, to get the 3D shape ear images, they exploited the Shape-From-Shading (SFS) algorithm. More recently, 3D shape ear images have also been used in [35] by using block-wise statistics to improve ear recognition system. However, the computation complexity of 3D ear recognition system is still expensive and high costs [36], [37], which limits its use for real-time applications. We can conclude that all previous researches have proved that the multimodal biometric authentication provides increased system security, recognition accuracy, and compensates for the limited resources of unimodal biometric recognition systems [15], [28], [32], [38]. Multimodal biometric recognition systems are guaranteed from forgery and theft and can achieve higher security level than those of unimodal biometric recognition systems those can be cheated. It has been well proved that multimodal biometrics recognition systems can override the respective limitations of unimodal system. It is interesting to exploit these advantages of multimodal biometric recognition systems to develop robust, efficient, and effective multimodal human recognition systems.

The main goal of metric learning is to learn a valid distance from a similarity function of a given problem or from the training data. There are a lot of existing metric learning approaches based on non-convex and convex optimization techniques or multi-kernel classification. However, Mahalanobis distance is considered as the most used distance in metric learning research. Mahalanobis distance $M_{\mathbf{A}}$ between two samples $x_i,\ and\ x_j$ is known as the squared distance under Euclidean distance over the new mapping space, $M_{\mathbf{A}}\left(x_i, x_j\right) = (x_i - x_j)^T \mathbf{A}(x_i - x_j)$, where $\mathbf{A} = \mathbf{L}^T\mathbf{L}$, and $\mathbf{A}$ is a semi-positive definite. SVM is popular classifier utilizing Mahalanobis distance and it has two main composing methods: one-versus-one (1-v-1) and one-versus-rest (1-v-r) classification methods. The 1-v-1 SVM is known as the binary classification. However, for many applications, the multi-class classification with SVM is needed. Hence, some of the multi-class classification methods have been presented in [39]–[42]. The DAGSVM is considered as one of multi-class SVM classification method. It works as 1-v-1 method in the training phase and adopts a root binary Directed Acyclic Graph (DAG) in the testing phase [39], [43], [44]. Nevertheless, the 1-v-r SVM classifier take more time for the learning process and scale linearly with respect to the number of classes. In addition, some of them are sensitive to solving some applications, especially when the data feature dimension is high, or the size of the training data is large. Therefore, for non-linear SVMs those separate the data points by using a non-linear decision boundary, several kernel tricks have been proposed such as Polynomial, Quadratic, and RBF kernels. Among them, the RBF is the most popular one used for many applications and it achieves a good classification performance.

In our work, feature-level fusion, which demonstrates more details and information, is adopted. It leads to better recognition performance. The LBP [45] and HOG [46] features are used to represent the face and ear images. The LBP is known as a powerful texture descriptor for improving system performance and recognition accuracy when used together with HOG features [47]. Moreover, the proposed model is based on learning distance metric and kernel SVM, which can be easily integrated in a multimodal biometric recognition system to increase the system performance. However, selecting and weighting features are also always big challenging issues in biometrics applications. Selecting a discriminative features may be quite challenging, when the feature vectors $\mathbf{x} \in z^p$ are in a space of high dimension p. Hence, it is required to find good discriminant features with lower dimension to represent face and ear images. The DCA algorithm for this purpose in [16]. It was investigated to combine discriminative local features and reduce the dimensionality of the Mahalanobis data matrix $\mathbf{M}$ before learning the distance metric.

III. THE PROPOSED LEANING FRAMEWORK: LEARNING MAHALANOBIS DISTANCE VIA DAGSVM

The proposed framework has depends on LBP and HOG features for face and ear images representation. To select discriminative features with lower dimension for each image, DCA is exploited as a dimension reduction technique, and a feature fusion method that incorporates the class association into correlation analysis to reflect the class structure information is utilized. After the DCA is conducted on the whole face, ear, or multimodal datasets, $N-1$ correlation components are left where N refers to the number of subjects in each dataset. **Algorithm 1** explains the summary of the DCA feature fusion algorithm, and for more details, see [16].

Algorithm 1 DCA Local Feature Fusion Algorithm

```
Input: LBP($\mathbf{X}_1$), HOG ($\mathbf{X}_2$)
Output: Y.
1 Select the class center $\bar{\mathbf{X}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} \mathbf{x}_j^i$ and the data
    center $\bar{X} = \frac{1}{n}\sum_{i=1}^{c} n_i\bar{\mathbf{X}}_i$, $n = \sum_{i=1}^{c} n_i$.
2 Calculate the covariance matrix $\mathbf{C} = \mathbf{\Phi}^T\mathbf{\Phi}$,
    where $\mathbf{\Phi} = \left(\sqrt{n_1}\left(\bar{X}_1 - \bar{X}\right), \ldots, \sqrt{n_c}\left(\bar{X}_c - \bar{X}\right)\right) \in R^{d\times c}$.
3 Compute the Singular Value Decomposition (SVD) of
    $\mathbf{C} = \mathbf{U}\Lambda\mathbf{U}^T$, $\mathbf{\Lambda} = Diag\left(\lambda_1, \lambda_2, \ldots, \lambda_c\right)$, $(\lambda_1 \geq \lambda_2, \ldots, \geq \lambda_c)$; whereas $\lambda_i$ is the $i^{\text{th}}$ eigenvalue
    of $\mathbf{C}$.
4 Define the transformation matrix $\mathbf{P} = \mathbf{\Phi}\mathbf{U}_r\mathbf{\Lambda}_{r\times r}^{-\frac{1}{2}}$,
    where $r$ is the first largest non-zero eigenvalue and $\mathbf{U}_r$
    is the corresponding eigenvector.
5 Project $\mathbf{X}_1$ and $\mathbf{X}_2$ to $\mathbf{Z}_1 = \mathbf{P}^T\mathbf{X}_1$ and $\mathbf{Z}_2 = \mathbf{P}^T\mathbf{X}_2$.
6 Compute the between-set covariance matrix of the
    transformed feature set $\mathbf{S}_b = \mathbf{Z}_1\mathbf{Z}_2^T$.
7 Compute the SVD of $\mathbf{S}_b$; $\mathbf{S}_b = \mathbf{V}\Sigma\mathbf{V}^T$.
8 Define the transformation matrix $\mathbf{T}$: $\mathbf{T} = \mathbf{V}\mathbf{\Sigma}^{-\frac{1}{2}}$.
9 Transforme $\mathbf{Z}_1$ and $\mathbf{Z}_2$ to $\mathbf{X}_1' = \mathbf{T}^T\mathbf{Z}_1$ and $\mathbf{X}_2' = \mathbf{T}^T\mathbf{Z}_2$.
10 Combine data as $\mathbf{Y} = \mathbf{X}_1^{'} + \mathbf{X}_2^{'}$.
```

A hybrid LDM-DAGSVM model has been developed to investigate enough and robust multimodal biometric recognition based on face and ear images. Mahalanobis distance metric learning aims to search with a square matrix generated from the training set. The SVM has a better generalization ability than traditional classifiers such as the KNN using Euclidean distance. In the following Subsections, we will introduce each part of the proposed model.

A. SUPPORT VECTOR MACHINE (SVM)

Original SVM is designed for 2-class classification aiming at finding a hyper-plane to separate classes and maximize the margin; the margin refers to the distance between the hyper-plane and the closest points of both classes. Beside previous issues, minimization of the structural risk is adopted for SVM, referring to that a misclassification may not appear, when using the SVM classifier. Therefore, the SVM owns a better generalization ability than the traditional classifiers. Assume that the training dataset $\mathrm{T} = \{\mathbf{X}_i, Y_i\}, i = 1\ldots, n$, where $\mathbf{X}_i \in R^n$ is i^{th} input feature vector with output $Y_i \in \{0, 1\}$, which is the corresponding class label for $\mathbf{X}_i$. Therefore, the classification mapping is implemented considered as in Eq. (1).

$$Y_i = \boldsymbol{W}^T\boldsymbol{x}_i + b, \tag{1}$$

where $\mathbf{x}_i$ results from mapping of $\mathbf{X}_i$ according to $\mathbf{V} : R^n \rightarrow R^m$. This feature mapping that maps the input feature to a high dimensional space, nonlinearly [39]. $\mathbf{W} \in R^m$, and b is a bias term.

The optimization problem of SVM for separating classes is represented by the following Eq. (2) where ρ maximizes the geometric margin for the two classes.

$$\underset{\boldsymbol{W},b}{Max}\rho \quad s.t\ \frac{Y_i(\boldsymbol{W}^T\boldsymbol{x}_i + b)}{\|\boldsymbol{W}\|} \geq \rho, \quad \text{for all } i = 1\ldots, n. \tag{2}$$

Eq. (2) is equivalent to Eq. (3), and we can form it as follows.

$$\begin{aligned} &\underset{\boldsymbol{W},b}{Min}\frac{\|\boldsymbol{W}\|^2}{2} \\ &s.t\ Y_i\left(\boldsymbol{W}^T\boldsymbol{x}_i + b\right) \geq 1, \quad \text{for all } i = 1\ldots, n. \end{aligned} \tag{3}$$

where $\|\mathbf{W}\|$ is the norm of the normal vector weights of the hyper-plane. To obtain the constrained optimization problem, the primal Lagrange form is given as in Eq. (4).

$$\begin{aligned} L\left(\boldsymbol{W}, b, \alpha\right) = \frac{\|\boldsymbol{W}\|^2}{2} \\ -\sum_{i=1}^{n}\alpha_i\left[Y_i\left(\boldsymbol{W}^T\boldsymbol{V}\left(\boldsymbol{X}_i\right) + b\right) - 1\right], \end{aligned} \tag{4}$$

where $\alpha_i \geq 0$ are Lagrange multipliers, and $\boldsymbol{V}$ represents the mapping. Therefore, the decision function rule for classification for a testing set x_i is formed as in Eq. (5).

$$\begin{aligned} D\left(\boldsymbol{x}_i\right) &= sign(\boldsymbol{W}_0^T\boldsymbol{x}_i + b_0) \\ &= sign\sum_{i=1}^{n}\left[\alpha_i Y_i K\left(\boldsymbol{x}, \boldsymbol{x}_i\right) + b_0\right], \end{aligned} \tag{5}$$

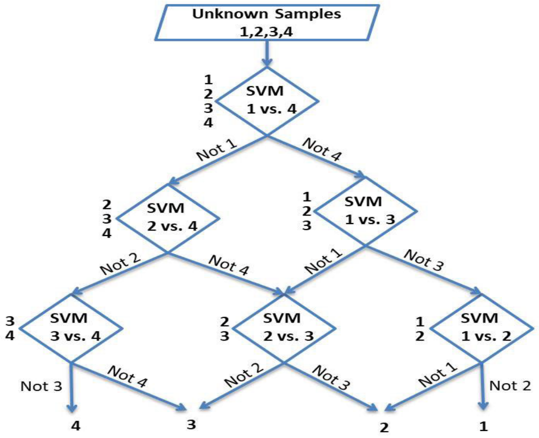


FIGURE 1. Decision diagram of DAGSVM for 4 classes labeled 1, 2, 3 and 4.

in which $K(x, x_i)$ is the kernel function [39] that represents any function satisfied Mercer's conditions [48]. In this work, the most common kernel function, called Gaussian RBF kernel, is adopted for the SVM classifier as follows in Eq. (6).

$$K(x, x_i) = e^{-\left(\frac{\|x-x_i\|^2}{2\sigma^2}\right)}, \tag{6}$$

where σ is a standard deviation of the Gaussian distribution.

B. DIRECTED ACYCLIC GRAPH SUPPORT VECTOR MACHINE (DAGSVM)

In the graph theory, DAG is defined as a graph with directed edges and without cycle's connection among vertices; directed edges mean that the edges go only one way from a vertex to another vertex. The DAGSVM [49] and DAGKNN are adopted in a root binary decision DAG learning structure in [39], [43], [44]. However, DAGSVM has shown great success compared to DAGKNN, especially with RBF kernel functions. The training model of DAGSVM [49] is the same as the 1-v-1 model that can be solved by $n(n-1)/2$ binary SVM classifiers. On the other hand, for the testing model, DAGSVM uses a rooted binary DAG that has $n(n-1)/2$ internal nodes organized in a diamond shape and n leaves labeled by classes as shown in Figure 1.

Figure 1 illustrates a decision diagram of DAGSVM for a 4-class data. Suppose $\mathbf{X}$ is a test sample. The classification operation is applied by starting from the root node, and each node works as a binary SVM classifier for classes of $\mathbf{X}_i$ and $\mathbf{X}_j$. Then, we go to either left edge or right edge depending on the SVM output value. Finally, we go through a path, and reach a leaf node that indicates the predicted output class. There exist many advantages for using DAGSVMs, it is analyzed that it can be established for better generalization; and the DAGSVM testing time is less than 1-v-1 SVM, since each node is trainined only with the labeled classes of the node, as shown in Fig.1 at the left side for each diamond node.

C. MAHALANOBIS DISTANCE VIA SVM

For simplicity, this work considers learning Mahalanobis distance via SVM by using DAGSVM and RBF kernel functions. For n given training samples, the Mahalanobis distance $M_{\mathbf{A}}$ between $\mathbf{x}_i$ and $\mathbf{x}_j$ is defined as in Eq. (7) [50].

$$\begin{aligned} M_A\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) &= \sqrt{\left(\boldsymbol{x}_i - \boldsymbol{x}_j\right)^T \boldsymbol{A} \left(\boldsymbol{x}_i - \boldsymbol{x}_j\right)}, \\ s.t.\, \boldsymbol{A} &= \boldsymbol{L}^T \boldsymbol{L}, \end{aligned} \tag{7}$$

where $\mathbf{x_i}, \mathbf{x_j} \in R^d$ and $\mathbf{A}$ is a positive semi-definite matrix [50]. Furthermore, $\mathbf{L} \in \mathbf{R}^{d\times d}$, and if $\mathbf{A} = \mathbf{I}^{d\times d}$, then Eq. (7) produces the Euclidean distance metric.

To transfer the Mahalanobis distance for learning the SVM, Eq. (7) should be integrated to the kernel function in Eq. (6). Then, calculate the result of mapping for the kernel function that is considered in Eq. (8).

$$\begin{aligned} K_L\left(\boldsymbol{x}_i,\boldsymbol{x}_j\right) = K_{ij} &= e^{-(\boldsymbol{x}_i-\boldsymbol{x}_j)^T \boldsymbol{L}^T \boldsymbol{L}(\boldsymbol{x}_i-\boldsymbol{x}_j)}, \\ s.t.\boldsymbol{L} = \boldsymbol{I}^{d\times d}\big/\sigma \quad & and\ \boldsymbol{L}_0 = \frac{\boldsymbol{I}}{d}, \end{aligned} \tag{8}$$

where $\mathbf{I}$ is the identity matrix, and σ is the standard deviation of the Gaussian distribution.

Finally, the training data is divided into a training set T and a validation set V. The SVM parameters are trained on T, and the outcome of the SVM is evaluated on the validation set V. Therefore, the objective of $\mathbf{L}$ is to maximize the inter-class variations and minimize the intra-class variations, which refers to minimize the classification error ε_V for the validation set V.

$$\begin{aligned} &\mathbf{L} = argmin\left\{\frac{1}{|\mathrm{V}|}\sum\nolimits_{(\mathbf{x},y)\in V}\ [D(\mathbf{x}) = y]\right\}, \\ &s.t.\,[D(\boldsymbol{x}) = y] \in \{0, 1\}, \end{aligned} \tag{9}$$

where $[D(\mathbf{x}) =y] = 1$ if and only if $D(\mathbf{x}) = y$. According to SVM classifier decision of Eq. (5) that relies on α and. The parameters α and b are re-trained for every intermediate setting of L. Since the $sign(.)$ function in Eq. (5) is non-continuous, performing the minimization to find $\mathbf{L}$ is then non-trivial, and it can be performed on a smooth loss function $\varphi_{\mathrm{V}}(\mathbf{L})$ as follows:

$$\begin{aligned} \varphi_{\mathrm{V}}(L) &= \frac{1}{|\mathrm{V}|}\sum\nolimits_{(\mathbf{x},y)\in V} \mathrm{s}_a[yD(\mathbf{x})], \\ s.t.\, s_a(z) &= {}^{1}\!\big/_{(1+e^{az})}, \end{aligned} \tag{10}$$

where $\mathrm{s}_a(z)$ is the sigmoid function, and a is a parameter that adjusts the steepness of the learning curve. For example, if $a \gg 0$, the function φ_{V} will be identical to ε_{V}. Since φ_V is a continuous and differentiable function. So, we can compute the derivative of φ_{V} to L. To derivative Eq. (5) and find $\partial D/\partial L$ that relies on α, K, and b, we should apply the chain rule as:

$$\frac{\partial D}{\partial \boldsymbol{L}} = \frac{\partial D}{\partial \alpha}\frac{\partial \alpha}{\partial \boldsymbol{L}} + \frac{\partial D}{\partial K}\frac{\partial K}{\partial \boldsymbol{L}} + \frac{\partial D}{\partial b}\frac{\partial b}{\partial \boldsymbol{L}}, \tag{11}$$

where $\frac{\partial D}{\partial \alpha}$, $\frac{\partial D}{\partial K}$, $\frac{\partial K}{\partial \mathbf{L}}$, and $\frac{\partial D}{\partial b}$ are straight-forward, therefore, we can compute $\frac{\partial b}{\partial \mathbf{L}}$. and $\frac{\partial \alpha}{\partial \mathbf{L}}$ that can be derived from the matrix inverse rule as follows:

$$\frac{\partial(\alpha, b)}{\partial \mathbf{L}_{ij}} = -\mathbf{H}^{-1}\frac{\partial \mathbf{H}}{\partial \mathbf{L}_{ij}}[\alpha, b]^T, \quad s.t.\ [\alpha, b]^T = \mathbf{H}^{-1}[1\ldots1, 0]^T \tag{12}$$

where $\mathbf{H} = \begin{pmatrix} \bar{K} & \mathbf{y} \\ \mathbf{y}^T & \mathbf{0} \end{pmatrix}$ and $\bar{K}_{ij}=y_iy_jK(\mathbf{x}_i,\mathbf{x}_j)$. Based on previous issues, our evaluations are restricted to the binary classification case, and we convert multi-class classification problems to binary ones. Additionally, we apply the cross-validation technique for each binary SVM classifier to prevent the overfitting of classification. Furthermore, the implementation of the proposed model depends on a modification of the SVM software LIBSVM [51].

IV. DATABASES AND EXPERIMENTAL RESULTS

To evaluate the feasibility, and effectiveness of the proposed framework, extensive experiments are conducted on public and available face and ear datasets, and their fusion constructed as a multimodal dataset. For all experiments, facial images are cropped beside the ear images. In addition, the face and ear images are converted to gray-scale images and resized into 100 × 100 pixels. Moreover, fusion of LBP and HOG descriptors is adopted to represent face and ear images. The proposed framework depends on uniform LBP with a radius of 2 pixels, a neighborhood size of 8 and a block size of 8 × 8 without block overlapping. On the other side, for the HOG descriptor, we adopt a cell of size 8 × 8 pixels, and a block size of 16 × 16 pixels with 8 pixels overlapping. Furthermore, the DCA is exploited as a feature level fusion algorithm that performs an effective feature fusion process based on maximizing the pairwise correlation across the two feature vectors. Therefore, the proposed model exploits the DCA algorithm advantages for feature fusion to avoid the limitations of using face or ear images, separately.

This section includes dataset description, performance evaluation metrics, and experimental results of unimodal face and ear recognition. Finally, the proposed efficient multimodal biometric recognition based on face and ear images is implemented by LDM with kernel SVM.

A. DATASET DESCRIPTION

Different biometric traits are adopted to perform our experiments. Ear datasets include two available and challenging ones: AWE and AMI. On the other hand, Georgia, ORL and the challenging AR face datasets are adopted to represent face datasets. Moreover, virtual multimodal biometric traits are constructed by fusing ear with face datasets, such virtual multimodal datasets called MD1 to MD6. In the following points, we describe each dataset, separately.

- Mathematical Analysis of Images (AMI) ear dataset. AMI dataset [52] has been collected from 100 persons, and each person has 7 images, totally 700 images. AMI ear images are collected from teachers, students, and staff at Universidad de Las Palmas de Gran Canaria (ULPGC), Spain. All persons are in the age range of 19 to 65 years, and ear images are taken in an indoor environment, of which the resolution is 492 × 702 pixels.
- Annotated Web Ear (AWE) dataset [53]. It includes 1,000 images of 100 subjects; each subject has 10 images. All images are collected from the Internet with various degrees of variability and illumination, and with different image scales and rotations. Hence, AWE ear dataset is considered as one of the most challenging ear datasets.
- Olivetti Research Laboratory (ORL) face dataset [54]. ORL dataset includes 400 images, 10 face images for 40 individual subjects. Face images are acquired from Cambridge employees and students. This dataset is collected with no restrictions imposed on expression. In addition, most of the person images are captured at different times with different lighting conditions. All images have the same resolution of 92 × 112 pixels, and some subject face images have glasses.
- Georgia Tech (GT) faces dataset [55]. To provide more face images for training, Georgia Tech face dataset is used, which contains 750 images for 50 subjects, and every subject has 15 color images. Face images are collected with different scales and orientations. Furthermore, most of the face images are taken in two or three different sessions that refer to variations in different illumination conditions, appearances, and facial expressions (open or closed eyes, smiling or not smiling). The average size of the face images is 150 × 150 pixels.
- AR faces dataset [56, 57]. As more complicated face images, AR dataset has over 3000 RGB images with an average size 768 × 576 pixels for 126 subjects, 74 subjects are men and the rest are women. Everyone has participated in two sessions separated with two weeks and without restrictions on hair style, make-up, accessories, and scarves. Each subject also has at least 26 images including face images under different conditions such as face expressions, illumination and pose variations, accessories and partial occlusion like scarves, sunglasses, hairs, and beards. Therefore, AR face dataset is considered as a challenging face dataset. All images are cropped to 128 × 128 pixels.

Figure 2 illustrates sample face images cropped and the original ear images for each dataset. In addition, Table 1 explains the number of images and number of subjects for all datasets (face, ear, and virtual multimodal biometric datasets), which are used in our experiments. Multimodal datasets will be described later with more details.

B. EXPERIMENTAL SETTINGS AND EVALUATION METRICS

Combined LBP and HOG features are adopted as standard local features to represent face and ear images. Moreover, LDM via kernel SVM is investigated for ear classification based on an RBF-SVM classifier. In order to validate the proposed model, every dataset is divided randomly into two disjoint groups, one for training and another for testing. The

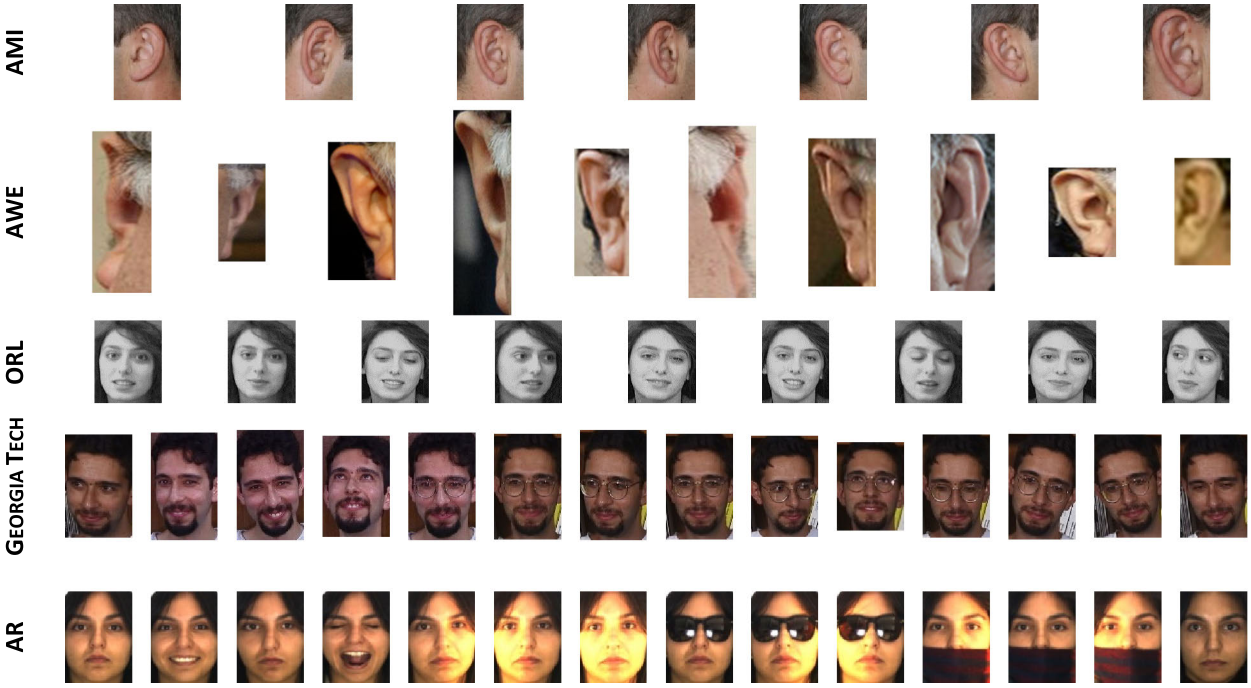


FIGURE 2. Sample images for ear and face databases.

TABLE 1. Face, ear and multimodal datasets.

Database	# of subjects	# of images
ORL face images	40	400
Georgia face images	50	750
AR face images	126	3276
AWE ear images	100	1000
AMI ear images	100	700
MD2: ORL + AWE images	40	800
MD2: ORL + AMI images	40	560
MD3: Georgia Tech + AWE images	50	1000
MD4: Georgia Tech + AMI images	50	700
MD5: AR + AWE images	100	2000
MD6: AR + AMI images	100	1400

5 fold cross-validation is adopted. The performance metrics are computed as the average Rank-1 recognition rates and standard deviations over all 5 folds for each experiment. In order to decrease the splitting influence of the training and testing data sets and to evaluate SVM training efficiency, all experiments are repeated for 10 times with randomly chosen training and testing data. The final recognition rate result is the average Rank-1 recognition accuracy of the 10 random experiments. The standard deviation is presented to assess the robustness for each face and ear datasets, and their fusion. Furthermore, to validate the proposed model performance, all experiments depend on two widely computed metrics: system accuracy (the average Rank-1 recognition rate) and the Receiver Operating Characteristic (ROC) curve. Model accuracy that represents the overall model performance on all available subjects can be formulated with Eq. (13), i.e., accuracy is the ratio of the number of correctly classified subjects divided by the test dataset size.

$$Accuracy\ (Acc) = {}^{(TP+TN)}\!\big/_{(TP+FP+TN+FN)} \quad (13)$$

where TP and FP refer to the true positive subjects and the false positive subjects, respectively. In addition, TN and FN mean the true negative and the false negative matched subjects. The proposed model adopts kernel RBF for SVM classifier to calculate the system accuracy that refers to the recognition rate. Besides, the proposed method depends on the ROC curve to evaluate and compare the model performance for each biometric trait and their fusion.

C. EFFICIENCY AND ROBUSTNESS OF DAGSVM TRAINING MODEL

This section explains the training efficiency of the DAGSVM model for our proposed approach improving learning Mahalanobis distance metric via kernel DAGSVM. In this subsection, we provide the proposed biometric model performance through 10 random repeated tests. Figure 3 shows the model accuracy variation during different repeating tests for AWE and AMI ear datasets. As shown in Figure 3, we can observe that the proposed model can achieve high performance within several repetitions times for the unimodal ear recognition. It gives around 95.50 % and 96.50 % on AWE and AMI ear datasets, respectively. The model performance on the AMI ear images is better than that on AWE images collected from the web as an unconstrained dataset.

Furthermore, for face recognition, the proposed model is evaluated on three standard and public face datasets (ORL, Georgia Tech and AR face databases) with 10 random

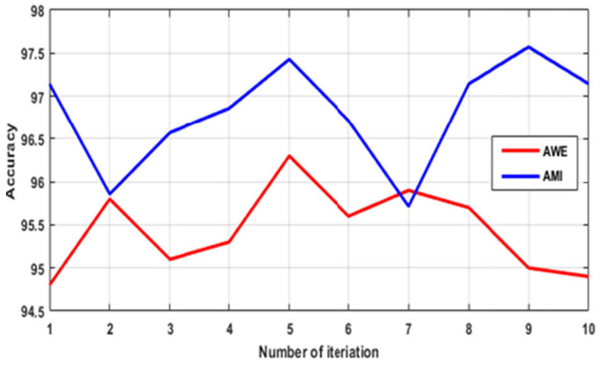


FIGURE 3. Model performance on AWE and AMI ear datasets through 10 repetitions.

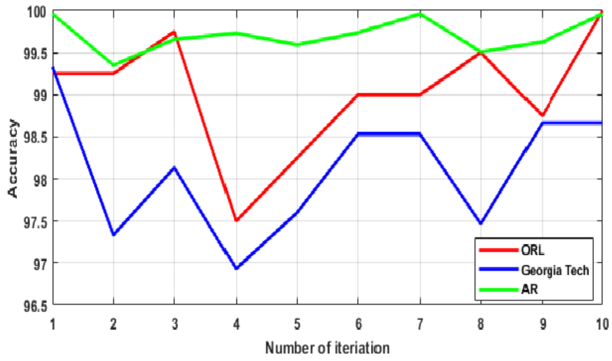


FIGURE 4. Model performance on ORL, Georgia Tech, and AR face datasets through 10 repetitions.

repetitions. The results prove the efficiency and robustness of DAGSVM training model as shown in Figure 4 for ORL, Georgia Tech and AR face datasets. Figures 3 and 4 prove the stability of the proposed model which achieves good results with several repetitions for unimodal face recognition system. Gratifyingly, the proposed face recognition model gives an accuracy around 99.00 %, 98.50 %, and 99.70 % for ORL, Georgia Tech, and AR face datasets, respectively. These results for unimodal ear and face recognition through several random repetitions reveal that the proposed model is stable more efficient than other classification models. Later, we will present the classifier performance through 10 repetitions for the proposed multimodal biometric recognition.

D. PERFORMANCE COMPARISON WITH OTHER MODELS

Shu *et al.* [58] and Kar *et al.* [59] and others [60]–[62] presented different models for face recognition. However, the performance of these models is not satisfactory as shown in Table 2 as in [58], [59], [62]. We can notice that Tables 2 and 3 obviously explain the proposed model achieves better performance than that of some recent state-of-the-art face and ear recognition models.

Moreover, the proposed model produces better recognition rates with low computation complexity compared to those of the models in [59], [63], [64]. These models adopt deep convolutional neural networks for improving face and ear image representation to achieve good accuracy. However, deep features are high-dimensional features which may lead to more computational complexity than that achieved with hand-crafted features (LBP and HOG) used in this work. Furthermore, most of the previous works on biometric recognition were based on traditional distance and classifiers [16], [53], [62], [64], [65].

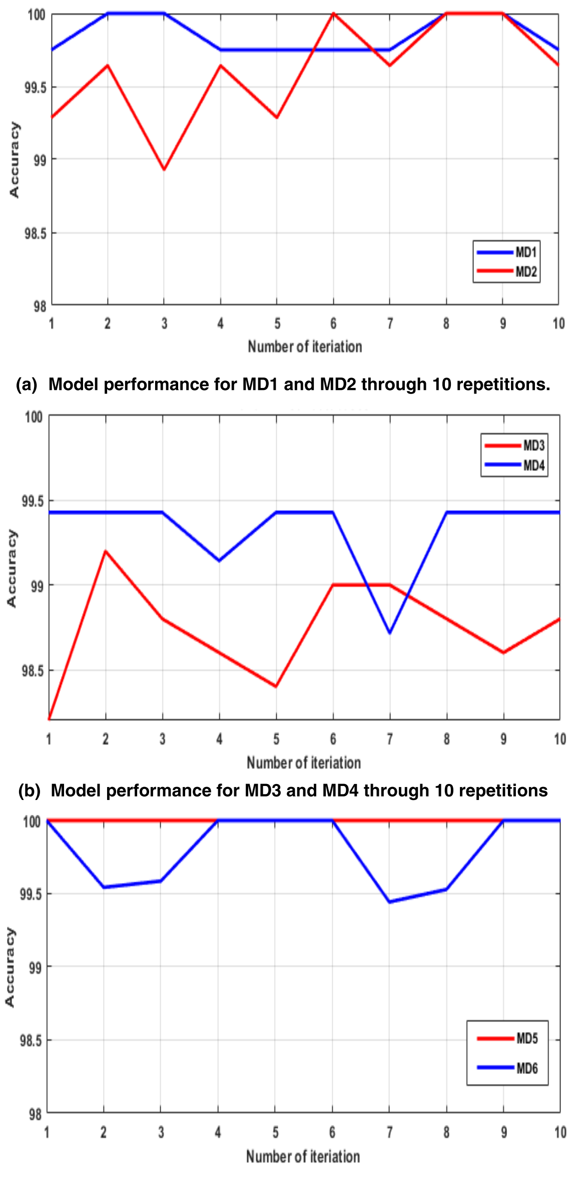
Multimodal biometric recognition has got more attention in the last decades, and it has been used in many intelligent applications such as immigration systems, access control systems, and surveillance systems. In order to improve the performance of a multimodal biometric recognition, the proposed model depends on combination of ear images with face

TABLE 2. Performance comparison in recognition rate (%) on AR dataset.

Methods			Recognition Rate
Shu *et al.* [58]	KNN	Gabor	64.28
	SRC		60.33
	CRC		70.92
	ProCRC		59.00
	SSRC-NW		92.14
Kar *et al.* [59]	CNN		96.33
	FaceNet		96.33
	MSLGFA		95.72
	TCPLRGF		96.35
Haghighat *et al.* [16]	KNN + HOG		85.14
	KNN + SURF		81.00
	KNN + Gabor		79.57
	KNN + HOG+SURF+Gabor		98.71
Junior *et al.* [60]	SVM + Curved Gabor		98.15
	SVM + Gabor		99.34
Khadhraoui *et al.* [61]	LGR+PuLBP		98.33
Huang *et al.* [31]	RSC + PCA		95.54
Xia *et al.* [62]	KNN		92.14
Proposed model	**LDM-DAGSVM**		**99.71±0.15**

TABLE 3. Performance comparison in recognition rate (%) on AWE dataset.

Methods			Recognition Rate
Emeršic *et al.* [53]	Chi-square distance	LBP	43.50±7.10
		BSIF	48.40±6.80
		LPQ	42.80±7.10
		RILPQ	43.30±9.40
		POEM	49.60±6.80
		HOG	43.90±7.90
		DSIFT	43.40±8.60
	Cosine distance	Gabor	39.80±7.10
Samuel *et al.* [63]	SVM	Alex Net	73.50
		VGG 16	81.50
		VGG 19	84.75
		ResNet	85.00
Zhang *et al.* [64]	Cosine distance	VGG-Face	87.50
Hassaballah *et al.* [65]	Chi-square distance	AECLBP	49.60 ± 3.40
Saeed *et al.* [66]	SVM	LGBP	54.20
Hansley *et al.* [67]	Sum fusion	CNN + HOG	90.60
Proposed model	**LDM-DAGSVM**	**LBP + HOG**	**95.44 ± 0.49**



(a) Model performance for MD1 and MD2 through 10 repetitions.

(b) Model performance for MD3 and MD4 through 10 repetitions

(c) Model performance for MD5 and MD6 through 10 repetitions

FIGURE 5. **Model performance generated datasets through 10 repetitions.**

images to construct robust and efficient multimodal feature vectors. To evaluate the proposed model for multimodal biometric recognition, we firstly constructed virtual multimodal datasets by using face and ear images and generated six virtual multimodal datasets called MD1 to MD6. However, the number of ear images is limited to 7∼10 images for each subject in the ear dataset. Therefore, the proposed takes the corresponding 7∼10 face images to the available ear images.

- MD1 (ORL+AWE) and MD2 (ORL+AMI). To build virtual multimodal datasets MD1 and MD2, the proposed method uses ORL face dataset with ear databases AWE and AMI. These two multimodal databases have 40 individuals; for each subject, ORL face images and corresponding ear images from AWE and AMI are randomly selected, respectively. Therefore, MD1 has 400 multimodal images in which

TABLE 4. **Average recognition rates (%) of the proposed model on different datasets.**

Databases	The proposed method (*LDM-DAGSVM*)
AWE ear images	95.44 ± 0.49
AMI ear images	96.82 ± 0.62
ORL face images	99.03 ± 0.73
Georgia face images	98.12 ± 0.76
AR face images	99.71 ± 0.15
MD1: ORL + AWE images	**99.85 ± 0.13**
MD2: ORL + AMI images	**99.61 ± 0.36**
MD3: Georgia Tech + AWE images	**98.74 ± 0.30**
MD4: Georgia Tech + AMI images	**99.33 ± 0.23**
MD5: AR + AWE images	**100**
MD6: AR + AMI images	**99.81 ± 0.25**

- MD1 is composed of face images from ORL face dataset and ear images from AWE ear dataset. It has 400 multimodal images. Each subject has 10 multi-traits composed of face and ear images. On the other side, MD2 is composed of face images from ORL dataset and ear images from AMI dataset. It has 280 multimodal images, each subject has 7 multi-traits.
- MD3 is composed of GT and AWE datasets, and MD4 is composed of GT and AMI datasets. MD3 contains 500 multi-images, and every subject has 10 multi-traits for 50 subjects. On the other hand, MD4 includes images for 50 individuals with 7 multi-traits for each one. Hence, the total number of images for MD4 is 350 images.
- MD5 is composed of AR and AWE datasets, and MD6 is composed of AR and AMI datasets. MD5 contains 1000 multi-traits for 100 individuals, while MD6 contains 700 multi-traits for 100 subjects.

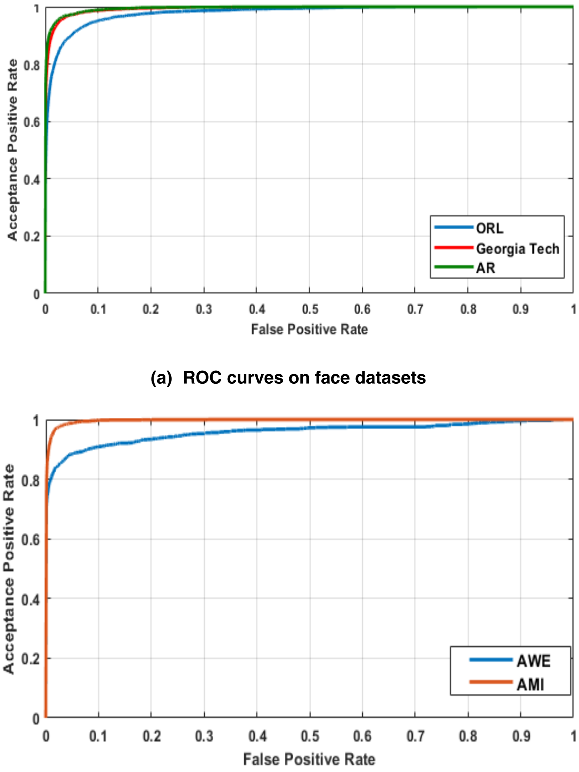
In order to evaluate the efficiency and robustness of the proposed DAGSVM model, experiments have been conducted on the virtual multimodal datasets MD1 to MD6. All experiments have been evaluated 10 times, whereas each multimodal dataset is proposed 5 fold cross-validation. Figures 5(a), 5(b), and 5(c) show the proposed model through 10 repetitions on MD1 to MD6.

Figure 5 reveals that means the proposed DAGSVM model is stable, robust, and efficient for multimodal biometric recognition. Furthermore, we can see that the results on MD5 and MD6 are the best as they have the largest sizes as shown in Table 1. This means that the proposed model can achieve excellent performance, especially when providing more images for training.

The experimental results prove that the proposed model can also improve the multimodal recognition based on the fusion of face and ear images compared to the case of using individual traits as shown in Table 4. It can be noticed from Table 2 that he proposed model gives an accuracy of 99.50 % for multimodal biometric recognition, while it gives

TABLE 5. Performance comparison in recognition rate (%) between different fusion and recognition frameworks.

Methods		Recognition rates
Huang *et al.* [31]	Face and ear images	99.04
Haghighat *et al.* [16]	Face and ear images	98.56
Fan *et al.* [68]	Face and ear images	93.40
Regouid *et al.* [13]	Ear, iris and ECG	99 ~ 100
Kisku *et al.* [10]	Face and ear images (IITK)	96.16
Kabir *et al.* [69]	Ear, palm-print and fingerprint images	99.47
Yang *et al.* [38]	Face and ear images	98.63
Rathore *et al.* [70]	Face and ear images (IITK)	99.36
	Face and ear images (UND-J2)	98.02
	Face and ear images (UND-E)	96.02
Proposed model (LDM-DAGSVM)	MD1: face and ear images	**99.85 ± 0.13**
	MD2: face and ear images	**99.61 ± 0.36**
	MD3: face and ear images	**98.74 ± 0.30**
	MD4: face and ear mages	**99.33 ± 0.23**
	MD5: face and ear images	**100**
	MD6: face and ear images	**99.81 ± 0.25**



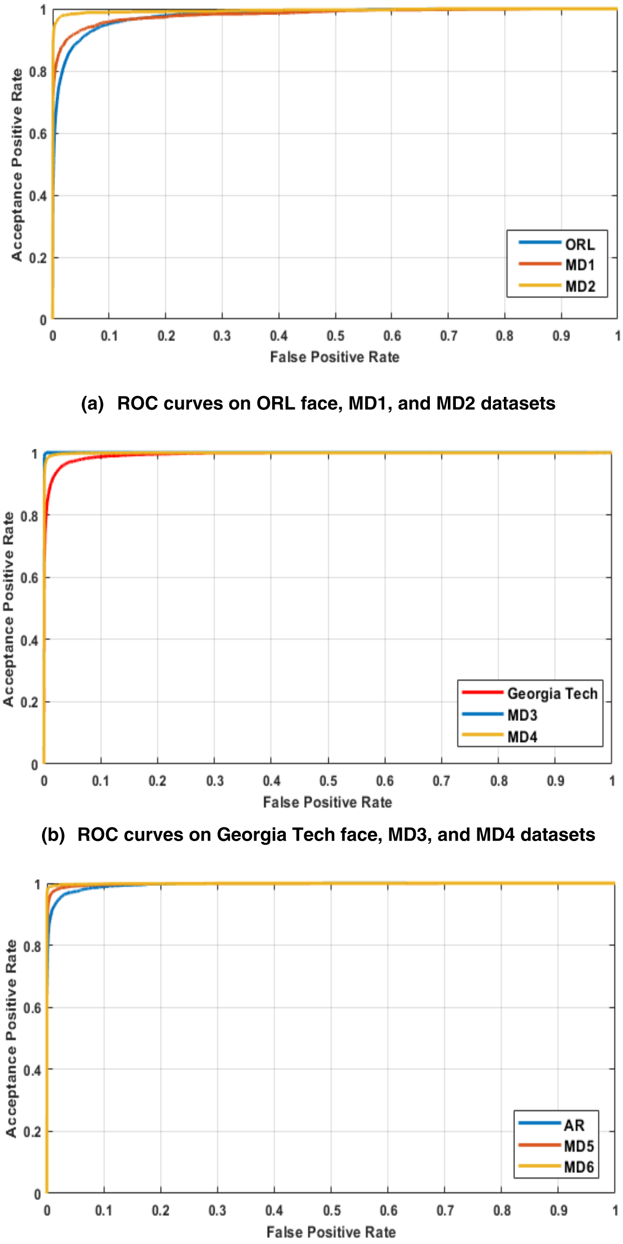
(a) ROC curves on face datasets

(b) ROC curves on ear datasets

FIGURE 6. System performance (ROC curves) for individual face and ear biometric systems.

accuracies of 99.00 % and 96.00 % for face and ear biometric recognition, respectively.

Moreover, we also introduce the ROC curves of the proposed model on MD1 to MD6 datasets as shown in Figures 7(a), 7(b), and 7(c). We can find that the proposed model gives better performance for multimodal biometric recognition compared to ear or face recognition only as shown in Table 4 and Figures 7(a), 7(b), and 7(c).



(a) ROC curves on ORL face, MD1, and MD2 datasets

(b) ROC curves on Georgia Tech face, MD3, and MD4 datasets

(c) ROC curves on AR face, MD5, and MD6 datasets

FIGURE 7. Performance of the proposed model on different datasets.

In addition, Figures 6(a) and 6(b) show the ROC curves for face and ear recognition. ROC curve refers to the model performance and presents the relation between the Acceptance Positive Rate (APR) and the False Positive Rate (FPR). Anyone can observe that the model performance on the AR face dataset is the best as it is large enough. In addition, the model performance on the AMI ear dataset outperforms AWE dataset as presented in Table 4 and Figure 6 (b). The AWE is considered as a challenging ear dataset [53, 63, 64, 65, 67],which contains ear images with high degrees of variability in pose, illumination and resolution as shown in Figure 2.

We have evaluated the proposed multimodal biometric recognition model and compared it with some state-of-the-art multimodal biometric models, especially those based on face and ear images as shown in Table 5.

Table 5 illustrates that the performance of the proposed multimodal biometric recognition model (LDM-DAGSVM) is superior to recent state-of-the-art multi-biometric models using face and ear images [10], [16], [31], [68], [70]. Moreover, we compared the proposed multimodal recognition model based on face and ear images to other multimodal biometric recognition models [13], [69] using different biometric traits. We can observe that the proposed model achieves competitive results compared to the other multi-biometric models using three biometric traits, and requiring more sensors to collect the data for three traits, not to mention the computation complexity of the traits.

V. CONCLUSION

The main motivation of the proposed multimodal biometric model includes two aspects: face and ear images representation, and human classification. Therefore, an efficient framework, based on a hybrid model of learning distance metric (LDM) and directed acyclic graph (DAG) support vector machine (SVM), has been proposed for a multimodal biometric recognition in this paper. Mahalanobis distance metric learning is used to seek a square matrix from the training set. Besides, kernel SVM is used to achieve better generalization ability than that of the traditional classifiers such as K-Nearest Neighbor (KNN) using Euclidean distance. Extensive experiments have been conducted on public and available face and ear datasets, and their fusions which are constructed as multimodal datasets. With experiments conducted on unimodal and multimodal datasets, the experimental results demonstrate that the proposed model is more effective than the other recent state-of-the-art human recognition models.

REFERENCES

[1] A. Jain, P. Flynn, and A. A. Ross, *Handbook of Biometrics*. New York, NY, USA: Springer, 2007.

[2] A. K. Jain and S. Z. Li, *Handbook of Face Recognition*. vol. 1. New York, NY, USA: Springer, 2011.

[3] K. Cao and A. K. Jain, "Hacking mobile phones using 2D printed fingerprints," Dept. Comput. Sci. Eng., Michigan State Univ., East Lansing, MI, USA, Tech. Rep. MSU-CSE-16-2, 2016.

[4] K.-H. Pong and K.-M. Lam, "Multi-resolution feature fusion for face recognition," *Pattern Recognit.*, vol. 47, no. 2, pp. 556–567, Feb. 2014.

[5] S. Umer, B. C. Dhara, and B. Chanda, "Face recognition using fusion of feature learning techniques," *Measurement*, vol. 146, pp. 43–54, Nov. 2019.

[6] P. P. Sarangi, B. S. P. Mishra, and S. Dehuri, "Fusion of PHOG and LDP local descriptors for kernel-based ear biometric recognition," *Multimedia Tools Appl.*, vol. 78, no. 8, pp. 9595–9623, Apr. 2019.

[7] F. Vallet, S. Essid, and J. Carrive, "A multimodal approach to speaker diarization on TV talk-shows," *IEEE Trans. Multimedia*, vol. 15, no. 3, pp. 509–520, Apr. 2013.

[8] J. Yang, J.-Y. Yang, D. Zhang, and J.-F. Lu, "Feature fusion: Parallel strategy vs. Serial strategy," *Pattern Recognit.*, vol. 36, no. 6, pp. 1369–1381, Jun. 2003.

[9] X. Pan, X. Xu, Y. Lu, and Y. Cao, "Feature fusion in multimodal recognition based on ear and profile face," *Proc. SPIE*, vol. 1, no. 1, p. 89, 2008.

[10] D. R. Kisku, P. Gupta, H. Mehrotra, and J. K. Sing, "Multimodal belief fusion for face and ear biometrics," *Intell. Inf. Manage.*, vol. 1, no. 3, pp. 166–171, 2009.

[11] I. Omara, X. Li, G. Xiao, K. Adil, and W. Zuo, "Discriminative local feature fusion for ear recognition problem," in *Proc. 8th Int. Conf. Biosci., Biochem. Bioinf. (ICBBB)*, 2018, pp. 139–145.

[12] E. E. Hansley, M. P. Segundo, and S. Sarkar, "Employing fusion of learned and handcrafted features for unconstrained ear recognition," *IET Biometrics*, vol. 7, no. 3, pp. 215–223, May 2018.

[13] M. Regouid, M. Touahria, M. Benouis, and N. Costen, "Multimodal biometric system for ECG, ear and iris recognition based on local descriptors," *Multimedia Tools Appl.*, vol. 78, no. 16, pp. 22509–22535, Aug. 2019.

[14] C. Ding and D. Tao, "Robust face recognition via multimodal deep face representation," *IEEE Trans. Multimedia*, vol. 17, no. 11, pp. 2049–2058, Nov. 2015.

[15] I. Omara, G. Xiao, M. Amrani, Z. Yan, and W. Zuo, "Deep features for efficient multi-biometric recognition with face and ear images," *Proc. SPIE*, vol. 10420, Jul. 2017, Art. no. 104200D.

[16] M. Haghighat, M. Abdel-Mottaleb, and W. Alhalabi, "Discriminant correlation analysis: Real-time feature level fusion for multimodal biometric recognition," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 9, pp. 1984–1996, Sep. 2016.

[17] L. Lakshmanan, "Efficient person authentication based on multi-level fusion of ear scores," *IET Biometrics*, vol. 2, no. 3, pp. 97–106, Sep. 2013.

[18] I. Omara, H. Zhang, F. Wang, A. Hagag, X. Li, and W. Zuo, "Metric learning with dynamically generated pairwise constraints for ear recognition," *Information*, vol. 9, no. 9, p. 215, Aug. 2018.

[19] F. Wang, W. Zuo, L. Zhang, D. Meng, and D. Zhang, "A kernel classification framework for metric learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 9, pp. 1950–1962, Sep. 2015.

[20] S. Banerjee and A. Chatterjee, "Image set based ear recognition using novel dictionary learning and classification scheme," *Eng. Appl. Artif. Intell.*, vol. 55, pp. 37–46, Oct. 2016.

[21] M. R. Alam, M. Bennamoun, R. Togneri, and F. Sohel, "A joint deep Boltzmann machine (jDBM) model for person identification using mobile phone data," *IEEE Trans. Multimedia*, vol. 19, no. 2, pp. 317–326, Feb. 2017.

[22] S. Xiang, F. Nie, and C. Zhang, "Learning a Mahalanobis distance metric for data clustering and classification," *Pattern Recognit.*, vol. 41, no. 12, pp. 3600–3612, Dec. 2008.

[23] I. Omara, A. Hagag, and W. Zuo, "Learning LogDet divergence for ear recognition," in *Proc. 2nd Int. Conf. Biometric Eng. Appl. (ICBEA)*, 2018, pp. 69–73.

[24] J. Lu, J. Hu, and Y.-P. Tan, "Discriminative deep metric learning for face and kinship verification," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4269–4282, Sep. 2017.

[25] H. Méndez-Vázquez, "Metric learning in the dissimilarity space to improve low-resolution face recognition," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, vol. 10125. Lima, Peru: Springer, 2017, p. 217.

[26] V. E. Liong, J. Lu, and Y. Ge, "Regularized local metric learning for person re-identification," *Pattern Recognit. Lett.*, vol. 68, pp. 288–296, Dec. 2015.

[27] A. V. Iannarelli, *Ear Identification* (Forensic Identification Series). Paramount, CA, USA: Paramont Publishing Company, 1989.

[28] K. Chang, K. W. Bowyer, S. Sarkar, and B. Victor, "Comparison and combination of ear and face images in appearance-based biometrics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 9, pp. 1160–1165, Sep. 2003.

[29] A. Ross and A. K. Jain, "Multimodal biometrics: An overview," in *Proc. Eur. Signal Process. Conf.*, Sep. 2004, pp. 1221–1224.

[30] H. Nejati, Z. Li, T. Sim, E. Martinez-Marroquin, and D. Guo, "Wonder ears: Identification of identical twins from ear images," in *Proc. Int. Conf. Pattern Recognit.*, Nov. 2012, pp. 1201–1204.

[31] Z. Huang, Y. Liu, C. Li, M. Yang, and L. Chen, "A robust face and ear based multimodal biometric system using sparse representation," *Pattern Recognit.*, vol. 46, no. 8, pp. 2156–2168, Aug. 2013.

[32] M. M. Monwar and M. L. Gavrilova, "Multimodal biometric system using rank-level fusion approach," *IEEE Trans. Syst., Man, Cybern. B. Cybern.*, vol. 39, no. 4, pp. 867–878, Aug. 2009.

[33] Y. Li, W. Yuan, H. Sang, and X. Li, "Combination recognition of face and ear based on two-dimensional Fisher linear discriminant," in *Proc. IEEE 4th Int. Conf. Softw. Eng. Service Sci.*, May 2013, pp. 922–925.

[34] M. H. Mahoor, S. Cadavid, and M. Abdel, "Multi-modal ear and face modeling and recognition," in *Proc. 16th IEEE Int. Conf. Image Process. (ICIP)*, Nov. 2009, pp. 4137–4140.

[35] L. Zhang, L. Li, H. Li, and M. Yang, "3D ear identification using block-wise statistics-based features and LC-KSVD," *IEEE Trans. Multimedia*, vol. 18, no. 8, pp. 1531–1541, Aug. 2016.

[36] H. Chen and B. Bhanu, "Human ear recognition in 3D," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 718–737, Apr. 2007.

[37] Y. Liu, B. Zhang, and D. Zhang, "Ear-parotic face angle: A unique feature for 3D ear recognition," *Pattern Recognit. Lett.*, vol. 53, pp. 9–15, Feb. 2015.

[38] L. Yuan, Z.-C. Mu, and X.-N. Xu, "Multimodal recognition based on face and ear," in *Proc. Int. Conf. Wavelet Anal. Pattern Recognit.*, Nov. 2007, pp. 1203–1207.

[39] R. Debnath, N. Takahide, and H. Takahashi, "A decision based one-against-one method for multi-class support vector machine," *Pattern Anal. Appl.*, vol. 7, no. 2, pp. 164–175, Jul. 2004.

[40] K.-B. Duan and S. S. Keerthi, "Which is the best multiclass SVM method? An empirical study," in *Proc. Int. Workshop Multiple Classifier Syst.* Berlin, Germany: Springer, 2005, pp. 278–285.

[41] G. Madzarov, D. Gjorgjevikj, and I. Chorbev, "A multi-class SVM classifier utilizing binary decision tree," *Infomatica*, vol. 33, no. 2, pp. 1–10, 2009.

[42] Y. Liu and Y. F. Zheng, "One-against-all multi-class SVM classification using reliability measures," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, vol. 2, Dec. 2005, pp. 849–854.

[43] H. Joutsijoki, M. Siermala, and M. Juhola, "Directed acyclic graph support vector machines in classification of benthic macroinvertebrate samples," *Artif. Intell. Rev.*, vol. 44, no. 2, pp. 215–233, Aug. 2015.

[44] H. Joutsijoki and M. Juhola, "DAGSVM vs. DAGKNN: An experimental case study with benthic macroinvertebrate dataset," in *Proc. Int. Workshop Mach. Learn. Data Mining Pattern Recognit.* Berlin, Germany: Springer, 2012, pp. 439–453.

[45] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[46] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jun. 2005, pp. 886–893.

[47] X. Wang, T. X. Han, and S. Yan, "An HOG-LBP human detector with partial occlusion handling," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep. 2009, pp. 32–39.

[48] H. Friedman, J. Friedman, and J. Friedman, "Another approach to polychotomous classification," Dept. Statist., Stanford Univ., Stanford, CA, USA, Tech. Rep., 1996. [Online]. Available: https://ci.nii.ac.jp/naid/10017594776/#cit

[49] J. C. Platt, N. Cristianini, and J. Shawe-Taylor, "Large margin dags for multiclass classification," in *Proc. Adv. Neural Inf. Process. Syst.*, 2000, pp. 547–553.

[50] R. De Maesschalck, D. Jouan-Rimbaud, and D. L. Massart, "The Mahalanobis distance," *Chemometrics Intell. Lab. Syst.*, vol. 50, no. 1, pp. 1–18, 2000.

[51] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, 2011.

[52] L. A. E. Gonzalez and L. Mazorra, "Face and body biometrics in the wild: Advances in the visible spectrum and beyond," Ph.D. dissertation, CTIM ULPGC, Madrid, Spain, 2008. [Online]. Available: http://www.ctim.es/research_works/-ami_ear_database/

[53] Ž. Emersic, V. Štruc, and P. Peer, "Ear recognition: More than a survey," *Neurocomputing*, vol. 255, pp. 26–39, Sep. 2017.

[54] (1994). *The ORL Database of Faces*. [Online]. Available: http://-cam-orl.co.uk/facedatabase.html

[55] L. Chen, H. Man, and A. V. Nefian, "Face recognition based on multi-class mapping of Fisher scores," *Pattern Recognit.*, vol. 38, no. 6, pp. 799–811, Jun. 2005.

[56] A. M. Martinez, "The AR face database," CVC, Ohio State Univ., Columbus, OH, USA, Tech. Rep. 24, 1998.

[57] A. M. Martinez and A. C. Kak, "PCA versus LDA," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 2, pp. 228–233, 2nd Quart., 2001.

[58] T. Shu, B. Zhang, and Y. Y. Tang, "Sparse supervised representation-based classifier for uncontrolled and imbalanced classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 8, pp. 2847–2856, Aug. 2020.

[59] A. Kar and P. P. G. Neogi, "Triangular coil pattern of local radius of gyration face for heterogeneous face recognition," *Appl. Intell.*, vol. 50, no. 3, pp. 698–716, 2020.

[60] E. G. L. Junior, L. H. S. Vogado, R. D. A. L. Rabelo, and C. J. P. Passarinho, "Curved Gabor projection entropy for face recognition," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2018, pp. 1–8.

[61] T. Khadhraoui, M. A. Borgi, F. Benzarti, C. B. Amar, and H. Amiri, "Local generic representation for patch uLBP-based face recognition with single training sample per subject," *Multimedia Tools Appl.*, vol. 77, no. 18, pp. 24203–24222, Sep. 2018.

[62] L. Xia, L. Wenhui, and S. Yixin, "Face recognition algorithm based on improved kernel sparse representation," in *Proc. 34rd Youth Academic Annu. Conf. Chin. Assoc. Autom. (YAC)*, Jun. 2019, pp. 654–659.

[63] S. Dodge, J. Mounsef, and L. Karam, "Unconstrained ear recognition using deep neural networks," *IET Biometrics*, vol. 7, no. 3, pp. 207–214, May 2018.

[64] Y. Zhang, Z. Mu, L. Yuan, and C. Yu, "Ear verification under uncontrolled conditions with convolutional neural networks," *IET Biometrics*, vol. 7, no. 3, pp. 185–198, May 2018.

[65] M. Hassaballah, H. A. Alshazly, and A. A. Ali, "Ear recognition using local binary patterns: A comparative experimental study," *Expert Syst. Appl.*, vol. 118, pp. 182–200, Mar. 2019.

[66] U. Saeed and M. M. Khan, "Combining ear-based traditional and soft biometrics for unconstrained ear recognition," *J. Electron. Imag.*, vol. 27, no. 5, 2018, Art. no. 051220.

[67] E. E. Hansley, "Identification of individuals from ears in real world conditions," Ph.D. dissertation, Dept. Comput. Sci. Eng., Univ. South Florida, Tampa, FL, USA, 2018.

[68] T.-Y. Fan, Z.-C. Mu, and R.-Y. Yang, "Multi-modality recognition of human face and ear based on deep learning," in *Proc. Int. Conf. Wavelet Anal. Pattern Recognit. (ICWAPR)*, Jul. 2017, pp. 38–42.

[69] W. Kabir, M. O. Ahmad, and M. N. S. Swamy, "Normalization and weighting techniques based on genuine-impostor score fusion in multi-biometric systems," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 8, pp. 1989–2000, Aug. 2018.

[70] R. Rathore, S. Prakash, and P. Gupta, "Efficient human recognition system using ear and profile face," in *Proc. IEEE 6th Int. Conf. Biometrics, Theory, Appl. Syst. (BTAS)*, Sep. 2013, pp. 1–6.

IBRAHIM OMARA received the bachelor's degree (Hons.) in mathematics and computer science and the master's degree (M.Sc.) in computer science from Menoufia University, Egypt, in July 2005 and May 2012, respectively, and the Ph.D. degree in computer science from the School of Computer Science and Technology, Harbin Institute of Technology (HIT), China, in October 2018. He worked as a Demonstrator and an Assistant Lecturer with the Faculty of Science, Menoufia University, from April 2007 to October 2018. He is currently an Instructor with the Department of Mathematics and Computer Science, Faculty of Science, Menoufia University. He also holds a postdoctoral position at the School of Computer Science and Technology, Huazhong University of Science and Technology. His current research interests include computer vision, machine learning (metric learning algorithms), and biometrics with an emphasis on ear and face biometrics.

AHMED HAGAG (Member, IEEE) received the B.Sc. degree (Hons.) in pure mathematics and computer science and the M.Sc. degree in computer science from the Faculty of Science, Menoufia University, Egypt, in 2008 and 2013, respectively, and the Ph.D. degree in computer science from the School of Computer Science and Technology, Harbin Institute of Technology, China, in 2017. In 2009, he joined the Faculty of Computers and Information Technology, Egyptian E-Learning University, Cairo, Egypt, as a Teaching Staff. He is currently a Lecturer with the Faculty of Computers and Artificial Intelligence, Benha University. He has authored several technical journal and conference papers. His current research interests include image processing and remote sensing image interpretation, especially compression, classification, and wireless communication.

SOULEYMAN CHAIB was born in Mostaganem, Algeria, in 1988. He received the B.S. and M.S. degrees in computer science from the University of Science and Technology of Oran—Mohamed Boudiaf, Bir El Djir, Algeria, in 2009 and 2011, respectively, and the Ph.D. degree from the School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China. His research interests include very high-resolution image classification and scene classification.

GUANGZHI MA is currently an Associate Professor with the School of Computer Science, Huazhong University of Science and Technology. His current research interests include medical image processing, artificial intelligence, and data mining.

FATHI E. ABD EL-SAMIE (Member, IEEE) received the B.Sc. (Hons.), M.Sc., and Ph.D. degrees from Menoufia University, Menouf, Egypt, in 1998, 2001, and 2005, respectively. He is currently a Professor with the Department of Electronics and Electrical Communications, Faculty of Electronic Engineering, Menoufia University. His current research interests include image enhancement, image restoration, image interpolation, super-resolution reconstruction of images, data hiding, multimedia communications, medical image processing, optical signal processing, and digital communications. He was a recipient of the Most Cited Paper Award from the *Digital Signal Processing* journal in 2008.

ENMIN SONG (Senior Member, IEEE) received the Ph.D. degree in electrical engineering and computer engineering from the Teesside University, U.K. After completing his Ph.D. degree, he was a Postdoctoral Researcher with the University of California at San Francisco (UCSF). He is currently a Professor with the School of Computer Science and Technology, Huazhong University of Science and Technology, China. His current research interests include medical image processing and medical image information analysis.