



(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 5, May 2025

⊕ www.ijirset.com 🖂 ijirset@gmail.com 🖄 +91-9940572462 🕓 +91 63819 07438





[www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405281|

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

Logeshkumar S, Balagurunathan D

P.G. Student, Department of Computer Applications, MPNMJ Engineering College, Chennimalai, Erode,

Tamil Nadu, India

Assistant Professor, Department of Computer Applications, MPNMJ Engineering College, Chennimalai, Erode,

Tamil Nadu, India

ABSTRACT: Metric learning has significantly improved machine learning applications such as face re-identification and image classification using K-Nearest Neighbour (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 Mahala Nobis 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 outperforms the state-of-the-art multimodal methods. The proposed model achieves five-fold classification accuracy around 99.85% for the face and ear images.

KEYWORDS: Biometrics, multimodal biometrics, face and ear images, Mahalanobis distance, metric learning, DAGSVM.

I. INTRODUCTION

Unimodal biometric authentication systems have received more attention in the last decades for intelligent applications such as Internet of Things (IoT), 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 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 be applied 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 to parallel





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405281|

fusion providing 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 classification problems and 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 attention in the last decades especially for machine learning applications such as classification applications [18]–[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 Neighbour (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 learns 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.





[www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405281|

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

Chang et al. [28] presented a multimodal biometric recognition system for human identi-cation 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 Gibraltar were used for extracting land marks 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 MA between two samples xi and xj is known as the squared distance under Euclidean distance over the new mapping space, MA xi xj (xi xj)T A(xi xj), where A LTL, and A is a semipositive denote. SVM is popular classic er utilizing Mahalanobis distance and it has two main com posing 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 adopt Saroo binary Directed Acyclic Graph (DAG) in the testing phase [39], [43], [44]. Nevertheless, the 1-v-r SVM classi er 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 datapoints 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.

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].

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





[www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405281|

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 $T = \{X_{i}, Y_{i}\}$ i = 1..., n. where X_{i} in $R \land n$ is i \land (th) input feature vector with output $Y_{i} \land 0, 1$ which is the corresponding class label for X_{i} .

Algorithm 1 DCA Local Feature Fusion Algorithm Input: LBP(X), HOG (X_{2}) Output: Y.

1 Select the class centre overline x i = 1 n i sum j = 1 to n_{i} x j ^ i and the data centre overline X = 1 n sum i = 1 to c n i overline X i ,n= sum i=1 ^ c n i I

2 Calculate C = Phi \wedge T * Phi where R \wedge (dc) Phi=(sqrt(n) 1 (overline x 1 - overline X),..., sqrt n c (overline x c - overline x)) in

3 Compute the Singular Value Decomposition (SVD) of $C = U^{(U \cap T)} A = Diag (1, 2, ..., \lambda\epsilon)$. ($\lambda t \ge lambda_{2}$; ..., $\ge lambda c$) whereas lambda_{f} is the i^(th) eigenvalue of C.

4 Define the transformation matrix P= Phi*U_{r} Lambda r* r $^-$ 1 2 where r is the first largest non-zero eigenvalue and U_{r} is the corresponding eigenvector.

5 Project X_{1} and x_{2} to $Z_{1} = P \land T \ast X_{1}$ and $Z_{2} = P \land T \ast X_{2}$

6 Compute the between-set covariance matrix of the transformed feature set $S_{b} = Z_{1} * Z_{2} \wedge T$

7 Compute the SVD of S b ;S b =V Sigma tilde $V \wedge T$.

8 Define the transformation matrix T: T=V Sigma . 1 Sigma $^{(-1/2)}$

9 Transformed Z_{1} and Z_{2} to $X_{1}' = T \land T * Z_{1} X_{2}' = T \land T * Z_{2}$

10 Combine data as $Y = X_{1} + X_{2}$

IV. DATABASSES AND EXPERIMENTAL RESULTS

To evaluate the feasibility, and effectiveness of the pro posed 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 grayscale 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 neighbourhood 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 multi modal 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





[www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405281|

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 colour images. Face images are collected with different scales and orientations. Further more, 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. Every one 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.

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 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



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





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|



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.

		Recognition
		Rate
Shu <i>et al.</i> [58]	KNN	64.28
	SRC	60.33
	CRC Gabor	70.92
	ProCRC	59.00
	SSRC-NW	92.14
Kar <i>et al.</i> [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 <i>et al.</i> [62]	KNN	92.14
Proposed model	LDM-DAGSVM	99.71±0.15

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

Т





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405281|

A. 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

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 pro- posed model depends on combination of ear images with face images to construct robust and efficient multimodal feature vectors. To evaluate the proposed model for multimodal bio- metric 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 $7 \sim 10$ face images to the available ear images





Model performance for MD1 and MD2 through 10 repetitions.

Model performance for MD3 and MD4 through 10 repetitions

Т





[www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405281|

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 arerandomly selected, respectively. Therefore, MD1 has 400 multimodal images in which

- MD2 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 com- posed 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 accuracies of 99.00 % and 96.00 % for face and ear biometric recognition, respectively.

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 <i>et al.</i> [10]	Face and ear images (IITK)	96.16
Kabir <i>et al.</i> [69]	Ear, palm-print and fingerprint images	99.47
Yang et al. [38]	Face and ear images	98.63
	Face and ear images (IITK)	99.36
	Face and ear images (UND-J2)	
	Face and ear images (UND-E)	96.02
	MD1: face and ear images	99.85 ± 0.13
	MD2: face and ear images	99.61 ± 0.36
	MD3: face and ear images	$\textbf{98.74} \pm \textbf{0.30}$
	MD4: face and ear mages	99.33 ± 0.23
	MD5: face and ear images	100
	MD6: face and ear images	99.81 ± 0.25





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|



Volume 14, Issue 5, May 2025 |DOI: 10.15680/IJIRSET.2025.1405281|

ROC curves on face datasets

ROC curves on ear datasets

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).



ROC curves on AR face, MD5, and MD6 datasets

FIGURE 8. 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

IJIRSET©2025





[www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405281|

We have evaluated the proposed multimodal biometric recognition model and compared it with some state-of-theart 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. Mahala Nobis 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 dis- trance. 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. K. Cao and A. K. Jain, "Hacking mobile phones using 2D printed finger- prints," Dept. Comput. Sci. Eng., Michigan State Univ., East Lansing, MI, USA, Tech. Rep. MSU-CSE-16-2, 2016.
- 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.
- S. Umer, B. C. Dhara, and B. Chanda, "Face recognition using fusion of feature learning techniques," Measurement, vol. 146, pp. 43–54, Nov. 2019.
- 4. 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.
- 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.
- 6. J. Yang, J.-Y. Yang, D. Zhang, and J.-F. Lu, "Feature fusion: Par- allel strategy vs. Serial strategy," Pattern Recognit., vol. 36, no. 6, pp. 1369–1381, Jun. 2003.
- X. Pan, X. Xu, Y. Lu, and Y. Cao, "Feature fusion in multimodal recogni- tion based on ear and profile face," Proc. SPIE, vol. 1, no. 1, p. 89, 2008.
- 8. 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.
- 9. 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.
- 10. 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.
- 11. M. Regouid, M. Touahria, M. Benouis, and N. Costen, "Multimodal bio- metric system for ECG, ear and iris recognition based on local descriptors," Multimedia Tools Appl., vol. 78, no. 16, pp. 22509–22535, Aug. 2019.
- M. Haghighat, M. Abdel-Mottaleb, and W. Alhalabi, "Discriminant cor- relation analysis: Real-time feature level fusion for multimodal biomet- ric recognition," IEEE Trans. Inf. Forensics Security, vol. 11, no. 9, pp. 1984– 1996, Sep. 2016.









INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY





www.ijirset.com