

A Novel Approach for Face Recognition Using Fused GMDH-Based Networks

El-Sayed El-Alfy¹, Zubair Baig², and Radwan Abdel-Aal¹

¹College of Computer Sciences and Engineering, King Fahd University of Petroleum and Minerals, KSA

²School of Science and Security Research Institute, Edith Cowan University, Australia

Abstract: This paper explores a novel approach for automatic human recognition from multi-view frontal facial images taken at different poses. The proposed computational model is based on fusion of the Group Method of Data Handling (GMDH) neural networks trained on different subsets of facial features and with different complexities. To demonstrate the effectiveness of this approach, the performance is evaluated and compared using eigen-decomposition for feature extraction and reduction with a variety of GMDH-based models. The experimental results show that high recognition rates, close to 98%, can be achieved with very low average false acceptance rates, less than 0.12%. Performance is further investigated on different feature set sizes and it is found that with smaller feature sets (as few as 8 features), the proposed GMDH-based models outperform other classifiers including those using radial-basis functions and support-vector machines. Additionally, the capability of the group method of data handling algorithm to select the most relevant features during the model construction makes it more attractive to build much simplified models of polynomial units.

Keywords: Face recognition, abductive machine learning, neural computing, GMDH-based ensemble learning.

Received May 30, 2015; accepted November 29, 2015

1. Introduction

Face recognition has received wide acceptance in many applications for automatic personal identification or verification [11, 40]. Examples of these applications include country border-control systems, criminal identification systems, intelligent surveillance cameras, restricted area access in airports and hospitals, identity management in smart work environments, organization and retrieval of digital photos, and access control to personal devices such as mobile phones and laptops [18].

Compared to many other biometrics, a face recognition system is less intrusive, uses contact-less hygienic devices, is more socially acceptable, and can provide highly reliable results for mass-scale human recognition. Moreover, it can be used both as a stand-alone system or to complement other biometric systems [13, 33, 39]. It applies image processing and pattern recognition techniques on images acquired using ordinary digital cameras. Though several approaches have been proposed in the literature, it is still an active area of research to improve the recognition accuracy and deal with special issues such as quality and flexibility of image acquisition [1, 6, 21, 26, 30, 37].

An automatic recognition system typically consists of three main stages: image preprocessing, feature extraction and reduction, and classification or template matching. Image preprocessing is an optional stage that aims at enhancing the quality of the image before going into further processing (it might involve image lighting adjustment, localization, re-sampling, noise reduction,

aligning, histogram equalization, etc.) [22]. This stage might address several other factors such as mitigating the effects of pose, orientation, occlusion, emotions, aging, etc. on subsequent processing stages. The next stage is feature extraction and reduction which analyzes the preprocessed images and computes a reduced set of facial features that can be used to discriminate between the enrolled individuals. This stage is central in the face recognition process and can distinguish among various face recognition systems. Both linear and nonlinear feature extraction methods were proposed and have been divided into two main categories: geometric feature based methods and appearance-based methods [1,8]. While the former category uses prominent facial landmarks (such as hair, nose, eyes, etc.) and relative distances between them for image feature extraction [35], the latter category provides a holistic approach that uses global features. Among the proposed and widely used appearance-based methods is eigenfaces [34]. Recently eigenfaces have been used with Gabor filters to efficiently recognize partially occluded faces [31]. In [5], a feature extraction method is proposed based on different types of ordinal measures derived from the Gabor images to handle inter-person similarity and intra-person variations in face images. In [27], the authors focused on web-scale face identification using a technique known as Linearly Approximated Sparse Representation-based Classification (LASRC).

Since the dimensionality of the feature vector can be very high which implies more storage and processing time, statistical feature reduction

techniques are commonly used to represent facial images by a smaller set of features. Among these techniques are Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) and kernel methods [7, 14, 24, 25, 32]. Several variations of LDA have been studied. For example, a weighted fuzzy LDA approach is adapted for better class estimation through membership degree-based descriptions of the training sample distribution [38]. When the dataset size is small, two different algorithms are given in [4] to extract the discriminative common vectors in the training set of the face database. Another approach using Support Vector Machines (SVMs) has been proposed to tackle the face recognition problem [12, 28]. In [20], a hybrid neural network solution for face recognition is presented. It combines a Self-Organizing Map (SOM) for dimensionality reduction with a convolutional neural network. Comparative studies of various techniques for face recognition can be found in [16, 29].

This paper proposes and investigates a new approach for face recognition. The idea of this approach is the fusion of various models created with the Group Method of Data Handling (GMDH) algorithm [10]. With its ability to iteratively build models for a given dataset during training, this technique offers several promising advantages: high classification accuracies with low false acceptance rates, inherent dimensionality reduction by selecting the most relevant features while learning the optimum model, and simpler developed models. Previously, GMDH networks have been successfully applied to solve many real-world problems, e.g., [15, 19]; yet they have not been attempted for face recognition. Fundamentally, the proposed approach can be used with a variety of face features. To prove the concept, we utilized principal component analysis for eigen-decomposition which is a widely-used approach for feature extraction and reduction. Using a benchmark dataset, the effectiveness of the proposed approach is evaluated and the results are reported in terms of recognition rate, false positive rate and false negative rate. It is compared with some other machine learning techniques as well.

The rest of this paper is organized as follows: section 2 describes the details of the face recognition process and outlines the proposed approach. Section 3 describes the adopted face database, performance measures, experimental settings, comparisons, results and discussions. Finally, section 4 summarizes the paper findings.

2. The Proposed Method

Given a captured facial image, it is required to correctly identify the person who has this face image. We briefly review the procedure for feature extraction in section 2.1, followed with a description of the computational recognition model in section 2.2.

2.1. Feature Extraction

In principle, any of the feature extraction methods suggested in the literature can be used with the proposed approach described in the following subsection. In this paper, we used PCA to compute a set of base images known as eigenfaces which are very popular in face recognition systems and are often used as a baseline when testing newly developed methods [14]. The eigenfaces capture and encode the variations among a collection of images in a training corpus. Mathematically, eigenfaces are the eigenvectors of the covariance matrix of the set of images in the training corpus. Each eigenface has the same size as the original images. A good approximation of each image can be obtained by a weighted linear sum of these eigenvectors with corresponding weights representing the image projections onto the various dimensions of the eigenspace.

Consider a training set TS of M gray-scale preprocessed images of frontal face views where each image is represented by $d \times k$ matrix of grayscale levels. The procedure for computing the eigenfaces and feature vectors can be summarized as follows [34]:

- Represent each image $I_i(x, y)$ by a column vector ζ_i of length $D = d \times k$ through a row-wise scanning of the matrix $I_i(x, y)$ from the top to the bottom and from left to right, i.e.,

$$\zeta_i = [I_i(1,1) \dots I_i(1,k) \dots I_i(n,k)]^T; i = 1, 2, \dots, M \quad (1)$$

Where T denotes matrix transpose.

- Compute the mean face vector, φ , as follows,

$$\varphi = \frac{1}{M} \sum_{i=1}^M \zeta_i \quad (2)$$

- Subtract the mean from each face vector to obtain mean-shifted images, i.e.,

$$\delta_i = \zeta_i - \varphi; i = 1, \dots, M \quad (3)$$

- Compute the M eigenvectors v_i and the corresponding eigenvalues λ_i of the matrix $A^T A$, where $A = [\delta_1 \delta_2 \dots \delta_M]$, such that,

$$A^T A v_i = \lambda_i v_i; i = 1, \dots, M \quad (4)$$

- Select and normalize the K most relevant eigenvectors, where $K \leq M$, which have the highest eigenvalues and compute the eigenfaces u_i , which are the eigenvectors of the covariance matrix AA^T

$$u_i = A v_i; u_i^T u_i = 1; i = 1, \dots, K \quad (5)$$

- Calculate the weights or projections for each mean-subtracted image x onto the K principal dimensions as follows,

$$w_i = u_i^T (\zeta_x - \varphi); i = 1, \dots, K \quad (6)$$

Once each image is represented by a feature vector containing the weights (a.k.a. projections) along the most significant eigenfaces, we can proceed to the next stage to build a computational model for the identification purpose. This will be explained in the following subsection.

2.2. Fusion of GMDH-Based Models

In this section, we describe two computational models based on self-organizing GMDH neural networks. The idea is to fuse the outputs of simple GMDH models with varying complexity and trained on different subsets of features; this can have the potential of boosting the recognition rate. The first approach is sketched in Figure 1, and is referred to as E_1 , where the output vectors of three GMDH models are combined at the decision level to generate a single output. Each GMDH-based model receives as input the features extracted from the captured face image and generates an output vector of real-valued values between 0 and 1. For n subjects, the length of each output vector is n , with one entry corresponding to each subject. The three vectors are merged into one augmented sorted vector of length $3n$. The final identification output is determined by applying the majority voting scheme on the identification numbers corresponding to the highest three values in the augmented sorted vector.

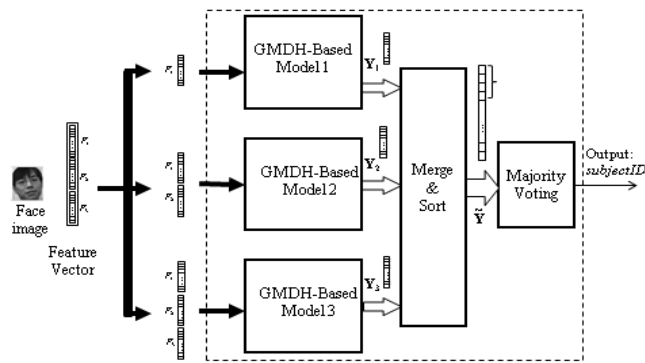


Figure 1. Ensemble model E_1 for recognizing n subjects by aggregating the decisions of three simple GMDH-based models through a majority voting scheme.

The second approach differs from the first approach in the way output vectors are merged. This approach is sketched in Figure 2 and is referred to as E_2 . The three output vectors from the simple GMDH-based models are averaged and the $argmax$ function is used to determine the identification number with the highest average value. Assume the elements in the vectors of the three GMDH-based models corresponding to the i -th subject are denoted y_{i1} , y_{i2} , and y_{i3} . Then, the corresponding entry in the average vector \bar{y}_i and the subject identification number are computed as follows,

$$\bar{y}_i = (y_{i1} + y_{i2} + y_{i3}) / 3; i = 1, \dots, n \quad (7)$$

$$subjectID = \underset{i=1, \dots, n}{argmax} \{ \bar{y}_i \} \quad (8)$$

For the sake of performance comparison, we also developed what we called a monolithic GMDH-based model to distinguish it from the ensemble models. Figure 3 shows a block diagram of this model. The first stage of this model is similar to the structure of one of the GMDH-based models adopted in Figure 1 and Figure 2. For n subjects, n binary recognizers are used with one associated with each subject. Each recognizer takes the feature vector as input and generates a real value y_i for $i = 1, 2, \dots, n$. The final identification number is determined using the $argmax$ function as follows,

$$subjectID = \underset{i=1, \dots, n}{argmax} \{ y_i \} \quad (9)$$

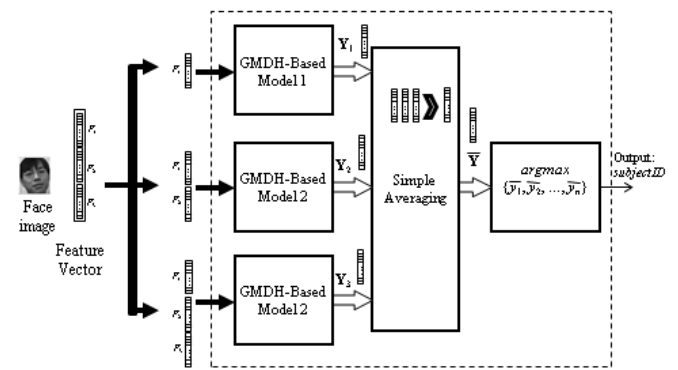


Figure 2. Ensemble model E_2 for recognizing n subjects using three simple GMDH-based models. The three output vectors from the three GMDH-based models are averaged and an $argmax$ function is applied to the resulting average vector to determine the final identification number.

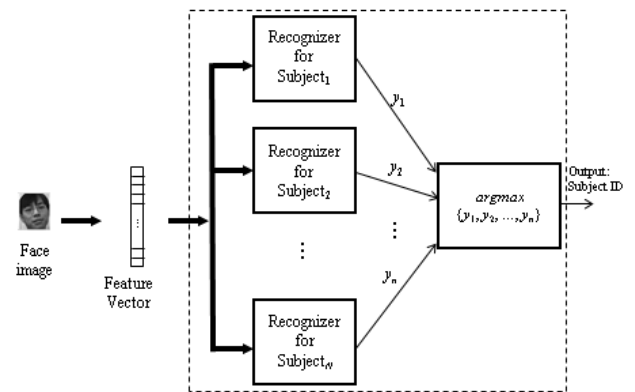


Figure 3. Monolithic abductive model for recognizing n subjects using n dedicated recognizers. The linear outputs from the n recognizers (y_1, y_2, \dots, y_n) are applied to an $argmax$ module to determine the subject identification number.

The construction of the recognizer for each subject is performed using the training dataset and a GMDH machine learning algorithm. The target variable is encoded to be 1 for the feature vectors of the associated subject and 0 otherwise. The adopted GMDH algorithm is known as Abductive Inductive Mechanism (AIM) [2]. It can automatically synthesize adequate models that embody the inherent structure of

complex and highly nonlinear systems. It builds layered feed-forward networks consisting of various types of polynomial functional elements. The network size, element types, connectivity, and coefficients for the optimum model are automatically determined using well-proven optimization criteria; thus reducing the need for user intervention when compared to traditional neural networks. This simplifies model development and considerably reduces the learning/development time and effort. Some examples of the functional elements are:

- *White elements*: the element output y is computed as a linear weighted sum of the element inputs, which are outputs of the previous layer, x_1, x_2, \dots, x_n as follows:

$$y = w_0 + \sum_{i=1}^n w_i x_i \tag{10}$$

Where w_0 is a constant and w_1, \dots, w_n are weights of corresponding inputs.

- *Single, double, and triple elements*: the element output y is computed from a third-degree polynomial with all possible cross-terms for one, two, and three inputs respectively; e.g.,

$$\text{Double: } y = w_0 + w_1 x_1 + w_2 x_2 + w_3 x_1^2 + w_4 x_2^2 + w_5 x_1 x_2 + w_6 x_1^3 + w_7 x_2^3 \tag{11}$$

The group method of data handling algorithm is a formalized paradigm for iterated polynomial regression capable of producing a high-degree polynomial model in effective predictors (input features). The process is evolutionary in nature, it starts with simple regression relationships between z -score normalized inputs and attempts to derive more accurate representations in subsequent iterations. To prevent exponential growth and limit model complexity, the algorithm selects only relationships having good predicting powers within each phase. Iteration is stopped when the new generation regression equations start to have poorer prediction performance than those of the previous generation, at which point the model starts to become overspecialized and therefore unlikely to perform well with new data.

AIM uses the Predicted Squared Error (PSE) criterion [3] for selection of inputs of each functional element and as a stopping criterion, to avoid model overfitting. This criterion minimizes the expected squared error that would be obtained when the network is used for predicting new data. The resulting *PSE* is described as follows:

$$PSE = FSE + CPM \left(\frac{2\kappa}{N} \right) \sigma_p^2 \tag{12}$$

where Fitting Squared Error (FSE) is the (averaged squared error) on the training data, Complexity Penalty Multiplier (CPM) is a selected by the user, κ is the

number of model coefficients, N is the number of samples in the training set, and σ_p^2 is a prior estimate for the variance of the error obtained with the unknown model. This estimate does not depend on the model being evaluated and is usually taken as half the variance of the dependent variable y [9]. As the model becomes more complex relative to the size of the training set, the second term increases linearly while the first term decreases. *PSE* goes through a minimum at the optimum model size that strikes a balance between accuracy and simplicity (exactness and generality). This trade-off can be optionally controlled using the *CPM* parameter. Larger values than 1 lead to the generation of simpler models that are less accurate but may generalize well with previously unseen data, while smaller values produce more complex networks that may overfit the training data and degrade actual prediction performance.

3. Experiments and Results

3.1. Dataset Description

In order to evaluate the effectiveness of the proposed face recognition approach, we used one of the publicly available benchmark datasets, the standard face database of the University of Manchester Institute of Science and Technology (UMIST) [11], in our experimental work. This database is very popular and has been used widely in the literature [11, 17, 23]. It consists of 575 gray-scale images of multiple views of 20 persons (subjects) taken at various angles from the left to the right. It covers several profiles and frontal poses for persons of different races, genders, and appearances. Persons are labeled using symbols a through t . The number of images per person ranges from 19 to 48. The original face images have varying sizes of approximately 220×220 in Portable Gray Map (PGM) format of 256 gray scales. The images are cropped into 112×92 sized arrays. Samples of these images are shown in Figure 4.



Figure 4. Sample of cropped images from the UMIST database.

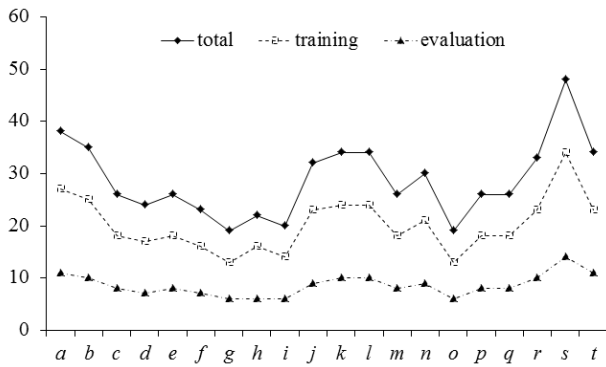


Figure 5. Distribution of the images for each of the 20 persons in the UMIST database between training and evaluation datasets (Subject#1 = a, Subject#2 = b, . . . , Subject#20 = t).

The images in the database are first processed using the eigenfaces mechanism, described in section 2.1, to represent each image by a feature vector composed of projections in the eigenspace. In our experiments, we set the maximum size of the feature vector to 64. Consequently, the dataset was randomized and then split into a training set of 403 images and an evaluation set of 172 images. The split was performed such that this ratio was satisfied for each of the 20 persons. Figure 5 shows the number of images selected for training and evaluation for each person in the dataset.

3.2. Performance Measures

The performance of different methods is evaluated and compared in terms of three measures: percentage recognition Error Rate (ER), percentage False Rejection Rate (FRR) and percentage False Acceptance Rate (FAR). The latter two metrics were calculated as the average values for the 20 persons. The definitions for these metrics are as follows:

$$ER = \left(\frac{N - \sum_{i=1}^n TP_i}{N} \right) \times 100\% \quad (13)$$

$$FRR = \left(\frac{1}{n} \sum_{i=1}^n \frac{FN_i}{AP_i} \right) \times 100\% \quad (14)$$

$$FAR = \left(\frac{1}{n} \sum_{i=1}^n \frac{FP_i}{AN_i} \right) \times 100\% \quad (15)$$

where n is the number of subjects ($n=20$), N is the total number of images in the evaluation set ($N=172$), AP_i is the number of images in the evaluation set belonging to the i -th subject (i.e., actual positive), AN_i is the number of images in the evaluation set belonging to all subjects other than the i -th subject (i.e., actual negative), TP_i is the number of images among the AP_i subset that were correctly classified as subject i (i.e., true positive), FN_i is the number of images among the AP_i subset that were classified as any subject other than subject i (i.e., false negative), TN_i is the number of images among the AN_i subset that were correctly classified as any subject other than subject i (i.e., true negative), and FP_i is the number of cases among the AN_i subset that were classified as

subject i (i.e., false positive).

3.3. Experiments

3.3.1. Monolithic GMDH-Based Models

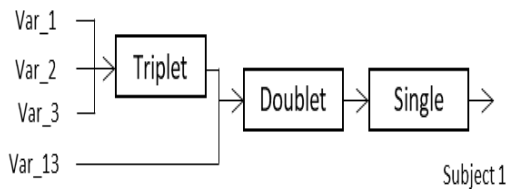
As the face identification problem is a multi-class problem, the first set of experiments involved training and evaluation of 20 monolithic binary classifier networks, one for each of the 20 subjects. These classifiers were trained on the 403 images of the training set and tested on the 172 images of the evaluation set. For training, each classifier output was set to 1 for cases corresponding to the associated subject and 0 otherwise; i.e., One-Versus-All (OVA) or winner-takes-all strategy. For evaluation, the *argmax* function was used to combine the outputs of all the 20 classifiers collectively. Thus, the class of the query subject image was determined to be that of the classifier giving the maximum output amongst the 20 classifiers. In all experiments, each of the 20 classifiers was individually optimized by selecting the *CPM* parameter that minimizes the absolute error for that classifier on the evaluation set.

We tested 5 different values for *CPM* (0.2, 0.5, 1.0, 2.0, and 5.0) and selected the best. The effect of the number of input features on the performance of the generated optimal models was investigated through the use of 8, 16, 32, and the full 64 features of the dataset. The results obtained for the percentage recognition error are illustrated in Table 1 for each of these cases which are referred to in the table as *A*, *B*, *C*, and *D*, respectively. These results indicate that the overall classification performance improves with increasing the number of features from 8 to 16, with over 50% reduction in the error rate. Increasing the number of features to 32 and 64 led to a slight degradation in the performance. This might be attributed to the relatively small size of the training dataset. The lowest recognition error rate is 4.65% which occurred when 16 features were used. However, if we change the criterion of correct classification to be ‘the correct class corresponds to one of the highest two outputs’ instead of being ‘the correct class corresponds to the highest output’, then the error rate drops to 2.33%.

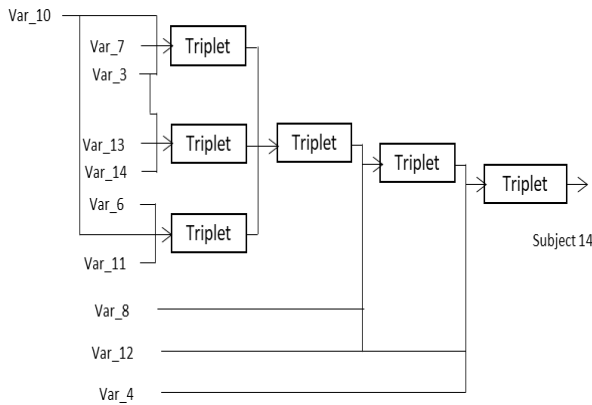
More experiments were conducted with the best model, denoted *B* in Table 1, which was created using 16 features. For this model, the 20 optimized classifiers were varied in complexity (from *CPM*=5 to *CPM*=0.2) and the input features for each classifier were automatically selected during training. It was found that the number of features selected for each classifier varied between 4 and 10, with an average of approximately 7 features selected per classifier. As an example, the simplest GMDH model structure was found to be for Subject#1 and the most complex model structure was found to be for Subject#14; see Figure 6.

Table 1. Recognition error rate for different monolithic and fused GMDH-based models (the best monolithic model and best fused model are found to be B and E_2).

Model	Input features	Error rate (%)	Description
A	1 to 8	9.88	Each of the models A, B, C and D is composed of 20 networks individually optimized by CPM
B	1 to 16	4.65	
C	1 to 32	5.81	
D	1 to 64	6.4	
E_1	1 to 64	3.49	Fusion of models B, C and D with majority voting
E_2	1 to 64	2.33	Fusion of models B, C and D by simple averaging of linear outputs



a) The simplest model (Subject#1).



b) The most complex model (Subject#14).

Figure 6. Model structures.

We also studied the impact of increasing the size of the classification problem at the optimum size of 16 features by increasing the number of subjects considered from 5 to 20 in steps of 5. As indicated by the results in Table 2, the method scales up well over the range of face classes available in the dataset. The high recognition error rates at the small sizes may be attributed to the dimensionality problem arising with the small training datasets.

Table 2. Performance of the best monolithic model, model B , as the size of the subject records is increased from 5 to 20 face classes.

Subjects	#Training records	#Evaluation records	Error rate (%)
1-5	105	44	6.82
1-10	187	78	7.69
1-15	287	121	4.13
1-20	403	172	4.65

Two other performance metrics are computed for the best monolithic model (model B): FRR and FAR . With standard processing, this model yields average values for FRR and FAR as 4.89% and 0.24%, respectively, as indicated in the first row of Table 3. In other words, an impostor is much less likely to be accepted as genuine as opposed to a genuine subject being classified as an

impostor.

Further attempts were made to enhance the classification accuracy and reduce the false alarms (a.k.a. reduce the FRR) as described in the next subsection.

Table 3. Performance comparison of best monolithic and best fused models (B and E_2 in Table 1, respectively).

Model	ER	FRR	FAR
B	4.65	4.89	0.24
E_2	2.33	2.21	0.12

3.3.2. Fused GMDH-Based Models

We investigated improving the recognition performance beyond that of the monolithic models through fusion of such models. Two methods were tested to combine the results from the best three monolithic models (B, C and D in Table 1). The first proposed model, denoted as E_1 in Table 1, performs fusion through majority voting among the three classification outcomes of the individual models. This model reduced the recognition error rate to 3.49% (as shown in Table 1). Further performance improvement was achieved with the second proposed fusion model, denoted as E_2 in Table 1. In this model, fusion was performed by simple averaging of the linear outputs of similar classifiers of the three models (i.e., before determining the output class from each model) then the $argmax$ is applied to determine the final output class. This reduced the recognition error rate further to 2.33%. Moreover, it helped reduce the FRR and FAR rates by over 50% as shown in the second row of Table 3. With FRR and FAR equal 2.21% and 0.12%, respectively, the best fused GMDH-based model tends to reject a genuine person more than accepting an impostor as genuine (which is of higher priority in security systems).

3.3.3. Comparisons and Discussions

For the sake of comparison with existing approaches, we considered six popular machine learning algorithms: k -nearest neighbor (k -NN), decision trees, rule based, Naive Bayes (NB), Radial-Basis Function (RBF) networks, and Support Vector Machines (SVMs). For k -nearest neighbor, we tested it for $k=1$ and $k=3$ with majority vote. For decision trees, we used the popular C4.5 algorithm. For rule-based, we applied the RIPPER algorithm classifier. Finally, for SVM, we applied an SVM with a polynomial kernel. The implementation details of these techniques are explained in [36] and the evaluation is performed for different number of input features. We calculated the recognition rate for each approach using 8, 16, 32 or 64 features and divided each value by the accuracy when using model E_2 , we call this ‘accuracy ratio’ metric, which is illustrated in Figure 7 for all methods with varying numbers of features. When the accuracy

ratio for a given method equals one, it implies that this method behaves as good as the best fused model. However, for accuracy ratio smaller than 1 (which is the case for most models), the best fusion model obtained is found to perform better. Otherwise, it performs worse.

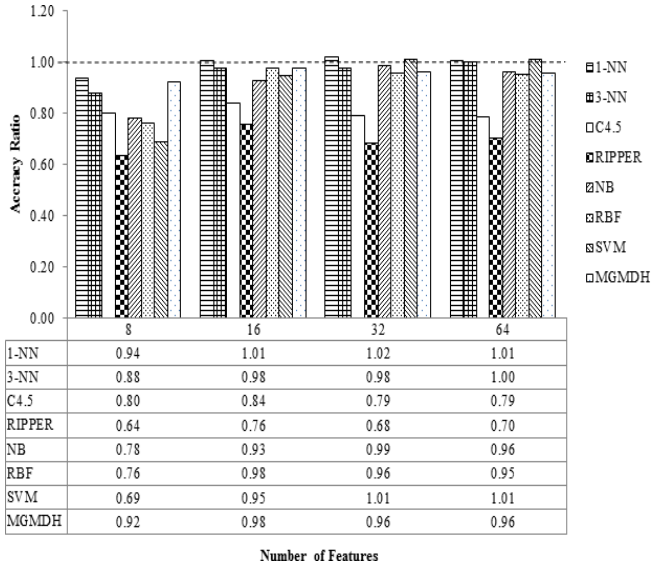


Figure 7. Comparing the accuracies of different methods for different number of features (where MGMDH refers to the monolithic GMDH models A, B, C and D for 8, 16, 32 and 64 features, respectively).

We also compared the *FRR* and *FAR* rates for various methods as shown in Table 4; we included the best monolithic and the best fused GMDH models for ease of reference. Based on these results, we made the following observations:

- As more features are used, the recognition rate enhances for most of the classifiers.
- The accuracy of *k*-NN with *k*=1 is the highest for the same number of features of 8, 16, and 32 features. But, when compared with the proposed best fused GMDH-based model, it is worse for 8 features and slightly better for 16 and 32 features.
- RIPPER has the worst recognition for all cases for the same number of features.
- For 8 and 16 features, the performance of the monolithic GMDH classifier was the second after the best performer (i.e., 1-NN) with a slight difference.
- For 8 features, the accuracy of none of the classifiers is higher than the proposed best fused GMDH-based model.
- Considering recognition rate, *FRR* and *FAR*, the best performance for SVM is slightly better than the proposed best fused GMDH-based model.

Overall, the proposed fused GMDH-based model is simpler and faster in terms of model building during training. It automatically selects fewer features and does not require retention of the original dataset after

building the model (as is the case with the 1-NN classifier, which is also known as the lazy classifier). Consequently, we can conclude that the proposed best fused GMDH-based model is either outperforming most classifiers or at least has the same recognition rate.

Table 4. Comparison of *FRR* and *FAR* for different methods when using 8, 16, 32 or 64 features (where *B* and *E*₂ refer to best monolithic GMDH and best fused GMDH in Table 1), respectively.

Model	8		16		32		64	
	<i>FRR</i>	<i>FAR</i>	<i>FRR</i>	<i>FAR</i>	<i>FRR</i>	<i>FAR</i>	<i>FRR</i>	<i>FAR</i>
1-NN	9.15	0.43	2.12	0.09	2.62	0.12	2.12	0.09
3-NN	15.77	0.73	5.39	0.25	1.74	0.10	2.63	0.13
C4.5	23.21	1.13	19.17	0.95	24.20	1.19	24.70	1.22
RIPPER	41.94	2.01	27.51	1.41	36.30	1.77	34.36	1.68
NB	26.29	1.26	10.38	0.49	4.34	0.19	5.66	0.31
RBF	26.99	1.35	5.38	0.24	7.25	0.33	8.30	0.36
SVM	39.17	1.72	10.04	0.41	1.67	0.08	1.08	0.08
<i>B</i>	-	-	4.89	0.24	-	-	-	-
<i>E</i> ₂	-	-	-	-	-	-	2.12	0.12

4. Conclusions

In this paper, we have presented a GMDH-based approach for human face recognition. Through empirical evaluation, a classification accuracy of a monolithic GMDH-based model was found to be 95.35% using only 16 features as input. The accuracy was further improved up to 97.67% through fusion of three GMDH-based models with the individual members trained separately using feature sets of different sizes. The results obtained from both the monolithic as well as the fused GMDH-based models were also compared with those obtained from seven other popular machine learning based classifiers in terms of recognition accuracies and false acceptance and rejection rates. The false rejection rate for the best monolithic model (with 16 input features) was found to be 4.89%, second only to the 1-NN classifier. This is reduced further using the proposed ensemble to 2.21%, which is on par with the classifier yielding the highest accuracy, i.e. 1-NN. Moreover, the ability of the GMDH algorithm to automatically select relevant features while constructing simpler models without having to retain the dataset, which is the case for nearest neighbor classifiers, gives the algorithm an edge in terms of the processing and storage capabilities needed. Other issues that can be further investigated in the future include fusion of multiple biometrics to enhance robustness for noisy data.

Acknowledgments

The authors would like to thank the support provided by King Fahd University of Petroleum and Minerals (KFUPM) through project No. RG1106-1&2 under the Intelligent Systems Research Group (ISRG).

References

- [1] Abate A., Nappi M., Riccio D., and Sabatino G., "2D and 3D Face Recognition: A survey," *Pattern Recognition Letters*, vol. 28, no. 14, pp. 1885-1906, 2007.
- [2] AbTech, *AIM User's Manual*, AbTech Corporation, 1990.
- [3] Barron A., Predicted Squared Error: A criterion for Automatic Model Selection, *Self-Organizing Methods in Modeling*, pp. 87-103, 1984.
- [4] Cevikalp H., Neamtu M., Wilkes M., and Barkana A., "Discriminative Common Vectors for Face Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 1, pp. 4-13, 2005.
- [5] Chai Z., Sun Z., Vazquez H., He R., and Tan T., "Gabor Ordinal Measures for Face Recognition," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 1, pp. 14-26, 2014.
- [6] Chakrabarty A., Jain H., and Chatterjee A., "Volterra Kernel Based Face Recognition Using Artificial Bee Colony Optimization," *Engineering Applications of Artificial Intelligence*, vol. 26, no. 3, pp. 1107-1114, 2013.
- [7] Chan L., Salleh S., Ting C., and Ariff A., "PCA and LDA-Based Face Verification Using Back-Propagation Neural Network," in *Proceedings of the 10th International Conference on Information Sciences, Signal Processing and their Applications*, Kuala Lumpur, pp. 728-732, 2010.
- [8] Chen W., Yuen P., Fang B., and Wang P., "Linear and Nonlinear Feature Extraction Approaches for Face Recognition," *Pattern Recognition, Machine Intelligence and Biometrics*, pp. 485-514, 2011.
- [9] Cottrell G. and Fleming M., "Face Recognition Using Unsupervised Feature Extraction," in *Proceedings of the International Neural Network Conference*, Paris, 1990.
- [10] Farlow S., The GMDH Algorithm, *Self-Organizing Methods in Modeling: GMDH Type Algorithms*, CRC Press, 1984.
- [11] Graham D. and Allinson N., Characterizing Virtual Egensignatures for General Purpose Face Recognition, *Face Recognition: from Theory to Applications*, Springer, 1998.
- [12] Guo G., Li S., and Chan K., "Support Vector Machines for Face Recognition," *Image and Vision Computing*, vol. 19, no. 9, pp. 631-638, 2001.
- [13] Kim Y., Shin K., Lee E., and Park K., "Multimodal Biometric System Based on the Recognition of Face and Both Irises," *International Journal of Advanced Robotic Systems*, vol. 9, no. 65, pp.1-6, 2012.
- [14] Kirby M. and Sirovich L., "Application of the Karhunen-Loeve Procedure for the Characterization of Human Faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 1, pp. 103-108, 1990.
- [15] Kondo T. and Ueno J., "Medical Image Diagnosis of Lung Cancer by A revised GMDH-Type Neural Network Using Various Kinds of Neuron," *Artificial Life and Robotics*, vol. 16, no. 3, pp. 301-306, 2011.
- [16] Kukreja S. and Gupta R., "Comparative Study of Different Face Recognition Techniques," in *Proceedings of the International Conference on Computational Intelligence and Communication Networks*, Gwalior, pp. 271-273, 2011.
- [17] Kumar D. and Shrutika J., "Harmony Search Algorithm for Feature Selection in Face Recognition," in *Proceedings of Communications in Computer and Information Science*, Pune, pp. 554-559, 2011.
- [18] Kurukulasooriya A. and Dharmarathne A., "Image Searching with Eigen faces and Facial Characteristics," in *Proceedings of Signal Processing, Image Processing and Pattern Recognition*, Jeju Island, pp. 215-224, 2011.
- [19] Lai I., Chang Y., Lee C., Chiou G., and Huang H., "Source Identification and Characterization of Atmospheric Polycyclic Aromatic Hydrocarbons Along the Southwestern Coastal Area of Taiwan-with a GMDH Approach," *Journal of Environmental Management*, vol. 115, pp. 60-68, 2013.
- [20] Lawrence S., Giles C., Tsoi A., and Back A., "Face Recognition: A Convolutional Neural-Network Approach," *IEEE Transactions on Neural Networks*, vol. 8, no. 1, pp. 98-113, 1997.
- [21] Liu Z., Zhao H., Pu J., and Wang H., "Face Recognition under Varying Illumination," *Neural Computing and Applications*, vol. 23, no. 1, pp. 133-139, 2013.
- [22] Lu Y., Zhou J., and Yu S., "A Survey of Face Detection Extraction and Recognition," *Computing and Informatics*, vol. 22, no. 2, pp. 163-195, 2012.
- [23] Luo M. and Zhang K., "A hybrid Approach Combining Extreme Learning Machine and Sparse Representation for Image Classification," *Engineering Applications of Artificial Intelligence*, vol. 27, pp. 228-235, 2014.
- [24] MageshKumar C., Thiagarajan R., Natarajan, S., Arulselv S., and Sainarayanan G., "Gabor Features and LDA Based Face Recognition with ANN Classifier," in *Proceedings of the International Conference on Emerging Trends in Electrical and Computer Technology*, Nagercoil, pp. 831-836, 2011.
- [25] Martinez A. and Kak A., "PCA Versus LDA," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 228-233, 2001.

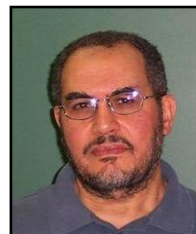
- [26] Oh B., Toh K., Teoh A., and Kim J., "Combining Local Face Image Features for Identity Verification," *Neurocomputing*, vol. 74, no. 16, pp. 2452-2463, 2011.
- [27] Ortiz E. and Becker B., "Face Recognition for Web-Scale Datasets," *Computer Vision and Image Understanding*, vol. 118, pp.153-170, 2014.
- [28] Phillips P., "Support Vector Machines Applied to Face Recognition," Technical Report NISTIR 6241, 1999.
- [29] Ruiz-del-Solar J. and Navarrete P., "Eigenspace-Based Face Recognition: A comparative Study of Different Approaches," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 35, no. 3, pp. 315-325, 2005.
- [30] Shah J., Sharif M., Raza M., and Azeem A., "A Survey: Linear and Nonlinear PCA Based Face Recognition Techniques," *The International Arab Journal of Information Technology*, vol. 10, no. 6, pp. 536-545, 2013.
- [31] Sharma M., Prakash S., and Gupta P., "An Efficient Partial Occluded Face Recognition System," *Neurocomputing*, vol. 116, pp. 231-241, 2013.
- [32] Shen L. and Bai L., "Gabor Feature Based Face Recognition Using Kernel Methods," in *Proceedings of the 6th IEEE International Conference on Automatic Face and Gesture Recognition*, Seoul, pp. 170-176, 2004.
- [33] Sun Y., Tang H., and Yin B., "The 3D Face Recognition Algorithm Fusing Multi-Geometry Features," *Acta Automatica Sinica*, vol. 34, no. 12, pp. 1483-1489, 2008.
- [34] Turk M. and Pentland A., "Face Recognition Using Eigenfaces," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Maui, pp. 586-591, 1991.
- [35] Wiskott, L., Fellous J., Kuiger N., and Von der Malsburg C., "Face Recognition by Elastic Bunch Graph Matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 775-779, 1997.
- [36] Witten I., Frank E., and Hall M., *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, 2011.
- [37] Xu Y., Li Z., Pan J., and Yang J., "Face Recognition Based on Fusion of Multi-Resolution Gabor Reatures," *Neural Computing and Applications*, vol. 23, no. 5, pp. 1251-1256, 2013.
- [38] Xue M., Liu W., and Liu X., "A novel Weighted Fuzzy LDA for Face Recognition Using the Genetic Algorithm," *Neural Computing and Applications*, vol. 22, no. 7-8, pp. 1531-1541, 2013.
- [39] Yu Q., Yin Y., Yang G., Ning Y., and Li Y., "Face and Gait Recognition Based on Semi-Supervised Learning," in *Proceedings of Chinese Conference on Pattern Recognition*, Beijing, pp. 284-291, 2012.
- [40] Zhao W., Chellappa R., and Phillips P., and Rosenfeld A., "Face Recognition: A Literature Survey," *ACM Computing Surveys*, vol. 35, no. 4, pp. 399-458, 2003.



El-Sayed El-Alfy is an Associate Professor and Coordinator of the Intelligent Systems Research Group, King Fahd University of Petroleum and Minerals, Saudi Arabia. His research areas include intelligent systems, information security, pattern recognition, digital forensic, network optimization, and traffic engineering. He has been actively involved in research projects and has 160+ refereed publications. He is a senior member of IEEE and on the editorial board of a number of reputable journals including IEEE Trans. Neural Networks and Learning Systems.



Zubair Baig is a member of the Security Research Institute and a Senior Lecturer of Cyber-Security in the School of Science at Edith Cowan University. He has 44 journal and conference articles and book chapters pertaining to Intelligent Network Security, Network Security/Performance Trade-off and Network Design and Optimization. His research interests are in the areas of cyber-security, artificial intelligence and optimization algorithms. He has served on numerous technical program committees of international conferences and has delivered a keynote talk on computer security.



Radwan Abdel-Aal received his BS in electrical engineering from Cairo University, Egypt, in 1972, his MS in aviation electronics from Cranfield University, UK in 1974, and his PhD from Strathclyde University, UK in 1983. Between 1985 and 2005, he was a research scientist at the Research Institute of King Fahd University of Petroleum & Minerals (KFUPM), Dhahran, Saudi Arabia. In 2005, he joined the Computer Engineering Department at KFUPM where he is currently a Professor. His research interests include nuclear physics instrumentation and machine learning and data mining applications.