IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received January 8, 2018, accepted February 27, 2018, date of publication April 11, 2018, date of current version April 25, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2813079

Classifiers Combination Techniques: A Comprehensive Review

MOHAMED MOHANDES, MOHAMED DERICHE, AND SALIHU O. ALIYU
Department of Electrical Engineering, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia

Corresponding author: Mohamed Mohandes (mohandes@kfupm.edu.sa)

This work was supported by DSR and CeGP, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia, under Grant GTEC1403.

ABSTRACT In critical applications, such as medical diagnosis, security related systems, and so on, the cost or risk of action taking based on incorrect classification can be very high. Hence, combining expert opinions before taking decision can substantially increase the reliability of such systems. Such pattern recognition systems base their final decision on evidence collected from different classifiers. Such evidence can be of data type, feature type, or classifier type. Common problems in pattern recognition, such as curse of dimensionality, and small sample data size, among others, have also prompted researchers into seeking new approaches for combining evidences. This paper presents a criteria-based framework for multi-classifiers combination techniques and their areas of applications. The criteria discussed here include levels of combination, types of thresholding, adaptiveness of the combination, and ensemble-based approaches. The strengths and weaknesses of each of these categories are discussed in details. Following this analysis, we provide our perspective on the outlook of this area of research and open problems. The lack of a well-formulated theoretical framework for analyzing the performance of combination techniques is shown to provide a fertile ground for further research. In addition to summarizing the existing work, this paper also updates and complements the latest developments in this area of research.

INDEX TERMS Pattern recognition, classifier combination, classifier ensembles, multi-classifiers, dimensionality reduction.

I. INTRODUCTION

The ability of humans in identifying fellow humans, objects, etc., lies in a learning process which starts right from the first day of birth. Similar learning processes have long been adopted to develop robust pattern recognition systems through the use of statistical, neural networks, genetic algorithms, etc. Such systems are important as they help in tasks involving large amount of data. Taking advantage of recent advances in computing power, developing such systems is of an immense importance in our daily life. Pattern recognition systems usually involve taking decisions or actions based on a given test pattern. The cost or risk of such an action varies from one application to another. The cost resulting from an action taken on a wrong DNA sequence identification is incomparable to the cost incurred for example in selling a sea bass fish for a salmon. Hence, the need to develop Pattern Recognition (PR) systems with confidence in their recognition accuracy is very crucial. To develop such PR systems, researchers often resolve to using evidences from more than one source (pattern classification systems) to achieve high recognition accuracy. This process of consulting more than one source or 'expert' before taking a final decision appears in the literature under several names such as multi-classifiers combination, multi-classifiers fusion, mixture of experts, ensemble based classification systems, etc.

While the idea of combining classifiers is not new, there is still a lot of scope for developing new combination approaches, types of features and classifiers used, and novel applications. In this section, we discuss some of the major surveys and reviews written on this topic. Xu *et al.* [1] discussed the different levels of classifiers combination based on the type of output (abstract, rank, or measurement) and associated problems. They discussed several combination techniques including: voting, Bayesian, and Dempster-Shafer (D-S) formulations, among others. These approaches were then used in identifying unconstrained handwritten numerals. The paper discussed in depth the suitability of different combination techniques for this application.

Kittler et al. developed a common theoretical framework for combining classifiers. In their work [2], they focused on ensembles which use distinct feature sets while in [3], both distinct and shared features were considered. Starting with the Bayesian decision rule, a mathematical framework was developed for the sum and the product rules. These two rules constitute the foundation for other rules including: max, min, median, and majority voting. The authors carried extensive experiments on handwritten digits recognition showing that the sum rule outperformed other rules despite being developed under the most restrictive assumptions. They also showed that the sum rule is much less affected by estimation errors; hence concurring that the developed theoretical framework is in agreement with previous experimental findings [2]–[4].

Duin and Tax conducted numerous experiments on different classifier combination rules [5]. The experiments include combination of same/different classifiers on same/different feature sets. The best performance was obtained when combining different feature sets with different individual classifiers. This is evident from the fact that the different feature sets/classifiers supply complementary information about the target classes, thus providing the best performance. In [6], a theoretical study comparing six classifier combination strategies was presented by Kuncheva. The techniques include average, min, max, median, majority vote, and oracle. The techniques were applied to a two-class problem assuming independent and identically distributed (IID) data. The classification error expressions were developed for all classifiers. Results show that the min and the max combination rules are the best for uniformly distributed posterior probabilities; likewise median and majority voting techniques have comparable performances for odd numbers of classifiers. However, the developed techniques were not extended for multi-class problems as well as non-IID cases.

Fumera et al. presented a theoretical and experimental analysis of multiple classifiers with a focus on linear combiners [7]. The authors particularly focused on simple and weighted averaging. Their analysis showed that the performance of linear combiners depends on the accuracy of the individual classifiers and the correlation between their outputs. They also showed that the weighted averaging rule outperforms simple averaging. In weighted averaging, finding the optimal weights for each classifier is still an open problem. Further research areas can involve the use of metaheuristic optimization approaches such as the Bat algorithm and other related techniques.

Polikar [8] discussed the idea of combining classifiers' output labels in comparison with combination using classifiers' continuous outputs as well as their application to ensemble based systems. In combining continuous outputs, the degrees of support given to each class, by the different classifiers, are used by the combination technique. This support is referred to as score in some literatures. The classifiers output labels are mainly combined using majority voting or its variants. The author primarily reviewed conditions for which ensemble based systems outperform individual classifiers. Different combination techniques were analyzed in details. In conclusion, the author stated that no single ensemble generation algorithm or combination rule is universally better than others. Tulyakov *et al.* [9] presented an overview of classifiers combination focusing on complexity, output type, and comparing ensembles vs. non-ensembles combinations. They also discussed the effects of retraining and issues related to locality in classifier combinations.

The paper in [10] reviewed classifiers combination approaches that operate after individual decisions are made. Under this category, the author discussed dynamic classifier selection, combination at decision, rank, and match score levels. The author also presented a universal model for converting any type of classifier output into a score so that the combination process at the score level can work more efficiently.

While there are several review papers in the literature, this paper presents a more updated survey including more recent approaches introduced during the past few years. Since the last major review paper [10], several novel techniques have been introduced. These include a signal strength based combining approach [11], a novel Bayes voting strategy [12], a modified weighted averaging technique using graph-theoretical clustering [13], a neural network based approach for training the combination rules [14], weighted features combination [15], and hierarchical fuzzy stack generalization [16], among others.

In this paper, we present a framework for categorizing existing combination approaches into subclasses sharing common characteristics. For each subclass, we discuss the advantages and disadvantages, computational complexity, and areas of applications.

The rest of this paper is organized as follows: section II presents the general framework for classifiers combination techniques, while section III discusses the most popular strategies used in multi-classifiers combination. Section IV concludes this paper with recommendations and future research directions.

II. GENERAL FRAMEWORK OF CLASSIFIERS COMBINATION

Typical classifiers combination algorithms start with a set of scores from individual classifiers, and produce a combined score for each class, followed by a final class label as shown in Fig. 1. Hence, the task is seen as a problem of finding a combination function which accepts N-dimensional score vectors from each of the M classifiers, then producing a single final classification score representing the selected class.

Similar to the notation used in [17], a given pattern Z is assigned to one of N possible classes $\{\omega_1, \omega_{2,} \ldots \omega_N\}$, using M classifiers. Each classifier receives as input a distinct measurement vector representing the given pattern. Let $\mathrm{x_k}$ be the measurement vector used by the $\mathrm{k^{th}}$ classifier. The probability density function, pdf, of the measurement vector is represented by $p(\mathbf{x}_k|\omega_n)$, while the prior probability

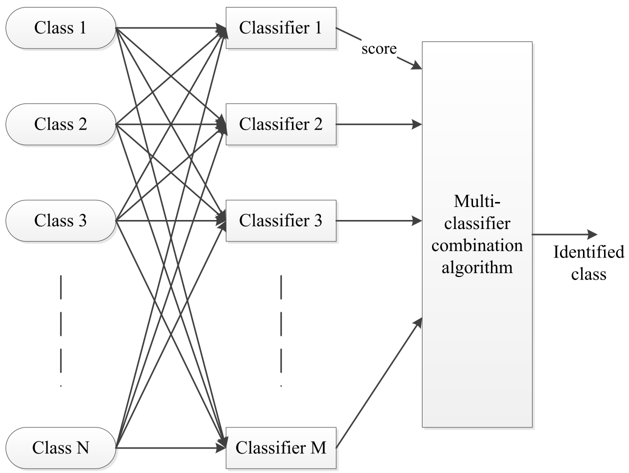

FIGURE 1. General framework for combining classifiers.

of occurrence of the class is denoted by $p(\omega_n)$. Using the Bayesian framework, and given the M observation vectors $\mathbf{x}_k$, k = 1,2,…,M, the pattern Z is assigned to class ω_j which produces the maximum aposteriori probability, i.e.

$$\begin{aligned} &assign \\ &\theta \rightarrow w_j \\ &if \\ &p(\theta = \omega_j | \mathbf{x}_1.\mathbf{x}_2., \ldots, \mathbf{x}_M) = \max_k p(\theta = \omega_k | \mathbf{x}_1, \ldots, \mathbf{x}_M) \end{aligned} \tag{1}$$

where θ is the label of pattern Z.

The expression above indicates that all information from the M classifiers was used in determining the final class label. Thus, it is necessary to compute the aposteriori probability of various hypotheses by considering all input observations to the classifiers concurrently. This requires the knowledge of the joint pdf, $p(\mathbf{x}_1, \ldots, \mathbf{x}_M | \theta = \omega_k)$ for all of the M classifiers which is obviously impractical. To simplify the above expression, it can be assumed that each classifier provides independently a decision support obtained from $\mathbf{x}_k$. To simplify the expression above, the aposteriori probability is re-written as:

$$p(\theta = \omega_j | \mathbf{x}_1, \ldots, \mathbf{x}_M) = \frac{p(\mathbf{x}_1, \ldots, \mathbf{x}_M | \theta = \omega_j) p(\omega_j)}{p(\mathbf{x}_1, \ldots, \mathbf{x}_M)} \tag{2}$$

where $p(\mathbf{x}_1, \ldots, \mathbf{x}_M)$ is the joint pdf of the observations independently of class label.

The M classifiers could be identical but use different feature sets as input, or use different parameter settings. Alternatively, the classifiers may be different in nature but use the same set of features as input. The important issue is that the individual classifiers should not make identical erroneous decisions on the same observation instances, i.e. they should provide complementary information. A successful combination of a set of individual classifiers information should improve the overall accuracy. This concept of classifiers combination has commonly been used as a means to improve accuracy, and has been widely applied in diverse applications including medical, military, economic, security and natural phenomena forecasting, among others.

III. CLASSIFIERS COMBINATION STRATEGIES

Over the past two decades, several classifiers combination techniques have been proposed in the literature. This section provides a detailed review of these techniques. The different techniques are grouped under classes sharing common characteristics. First, we discuss the different levels at which combination is performed. Such levels of combination include sensors, features, and decisions. We then expand on the concept of soft vs. hard combination techniques. Finally, we discuss how the combination techniques can be grouped as either adaptive or non-adaptive.

A. LEVELS OF CLASSIFIERS COMBINATION

Classifiers combination can be carried out at three different levels. These include: early combination which occurs at the sensor data level, or at the feature level, or a late combination at the decision level (see Fig. 2).

Sensor data level fusion involves combination of data collected from two or more sensors before feature selection technique is applied. Features level combination is simple and straightforward as it may simply involve basic concatenation of feature vectors with equal or different weights. The problem is that concatenation may result in high dimensional feature vectors. Such vectors may be transformed into low dimensional vectors using the popular PCA (Principal Component Analysis) technique with either linear or nonlinear kernels [18]. Sensor level fusion is free from possibility of high dimensional feature vectors as the feature extraction stage extracts only relevant information for classification stage. However, this is at the expense of the multiple sensors involved.

Most combination schemes operating at the decision level are based on one of three approaches: abstract, rank, and score. With the abstract-based approaches a single output label from each individual classifier is used as input to the combination scheme. With rank-based approaches, each classifier yields several labels ranked from the most likely to the least likely. This information is then used by the combination scheme to reach the final decision. The score (sometimes called measurement) based approach is the most informative technique [10]. Here, each classifier outputs the n best labels together with their confidence scores. The combination of scores can be achieved in a number of ways [10]. These include density based, transformation based, and classifier based score fusion. The score based combination approach can either use the Bayesian theory or the evidence theory, among others.

The evidence or D-S theory is a generalization of the Bayesian theory based on modeling uncertainty using the concept of belief functions [19]. The general expressions for the evidence theory (3) and the Bayesian theory (4) are given

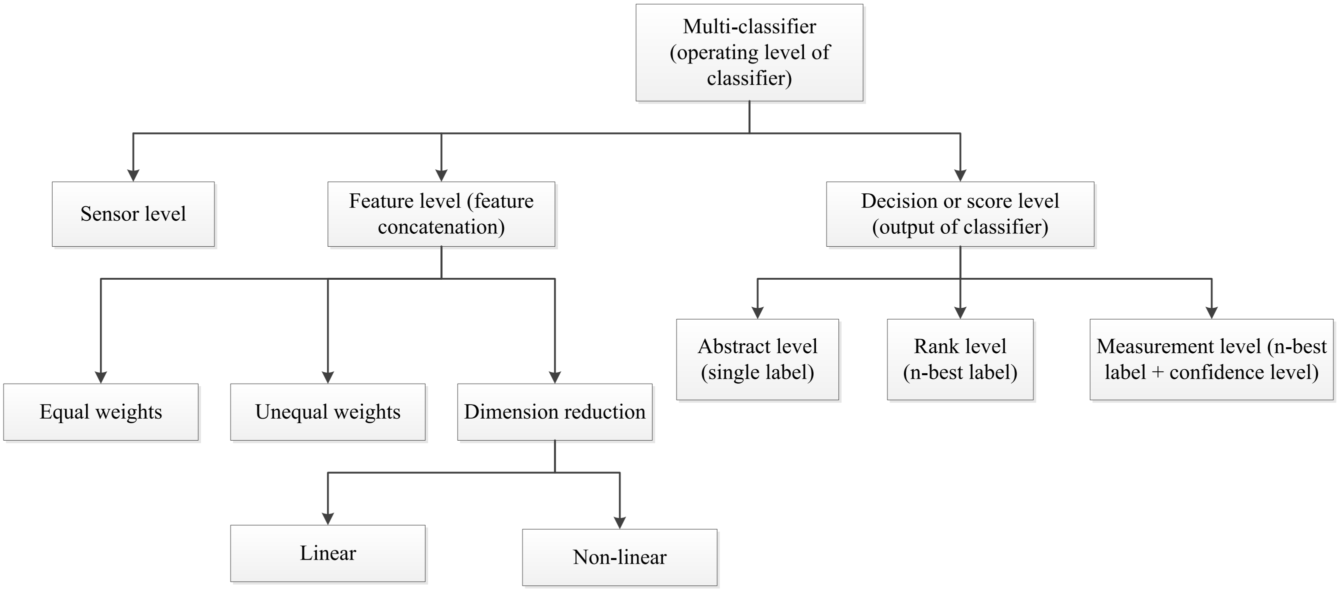

FIGURE 2. Different levels of classifiers combination.

respectively as:

$$p(A|C_1) + p(A|C_2) + \ldots + p(A|C_n) + \theta = 1 \quad (3)$$

$$p(A|C_1) + p(A|C_2) + \ldots + p(A|C_n) = 1 \quad (4)$$

Where θ measures uncertainty under the evidence framework. The evidence theory has been found useful in a wide range of applications where classifier uncertainty needs to be taken into account [19]–[26] at the expense of a high computational load [27].

The D-S theory is based on three main concepts: basic belief assignments, belief functions, and plausibility. The basic belief assignment (bba) assigns a value between 0 and 1 to all variables in a subset (A) where both the bba of the null set is 0 and the summation of bba's of all subsets should be equal to 1. Evidence is taken to be certain if the mass function $m(A) = 1$. However, belief function assigns a value [0, 1] to every non-empty subset B of A. Two bounds of an interval can be defined for every probability assignment. The D-S theory represents the lower bound by the belief function which is obtained from the sum of all basic belief assignments of the proper subsets B of A as in (5). The upper limit of the probability assignment is called plausibility. It is the sum of all probability assignments of subsets B that intersect the set of interest A as in (6) [19].

$$Bel(A) = \sum_{B \subset A} m(B) \quad (5)$$

$$Pl(A) = \sum_{B \cap A \neq \varphi} m(B) \quad (6)$$

where *Bel* represents the belief function and *Pl* represents the plausibility function. The overall combination rule is given in (7), where n is the number of available evidences.

$$m_n = \frac{\sum_{A_1 \cap A_2 \cap \ldots \cap A_n} m_1(A_1).m_2(A_2)\ldots m_n(A_n)}{1 - \sum_{A_1 \cap A_2 \cap \ldots \cap A_n = \varphi} m_1(A_1).m_2(A_2)\ldots m_n(A_n)} \quad (7)$$

As an example of applying the evidence theory in classification, Han *et al.* [27] used the D-S theory to fuse two classifiers, namely, BP and k-NN classifiers. It was tested on the Iris dataset in UCI. The dataset contains 150 samples with three classes and four features. Result obtained show that the performance of combined classifiers is better than individual classifiers. More evidences can be gathered if the features in the dataset used are heterogeneous. This was the case in [24] where different feature sets, selected based on different criteria, were used. The belief function theory led to two well-known combination rules: Dempster-Shafer and proportional conflict redistribution (PCR) theory. The approach was tested on online signature verification using a publicly available database. Results obtained showed the possibility of trade-off between recognition accuracy and reliability. Despite the successful use of D-S theory in many applications, some authors consider it a complex approach for classifiers combination [9].

In summary, we have seen that combination of classifiers can be achieved at different levels with no single strategy of being the best. However, we have also noticed that the tendency is towards fusing information after each of the classifiers has provided its labeling. This makes sense as the whole idea of fusion is to combine a bag of classifiers some of which may be weak. Note also that the strategy for fusion may either rely on a probabilistic approach, a learning approach, or an evidence based approach among others.

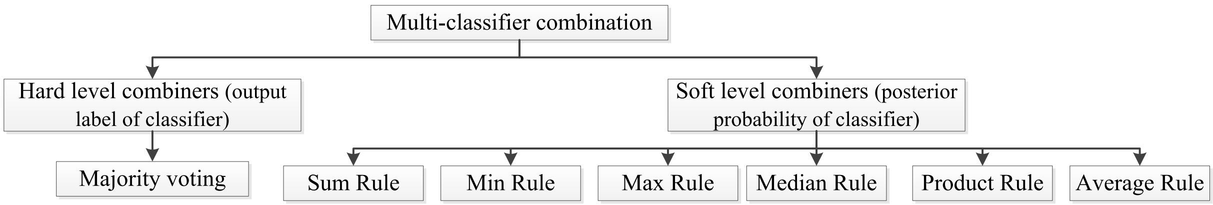

FIGURE 3. Hard and soft level classifiers combination schemes.

B. HARD AND SOFT LEVEL CLASSIFIERS COMBINATION

Another way to categorize combination algorithms is whether hard thresholding or soft scoring is used with each of the classifiers. Hard-level combination uses the output of the classifier after it is hard thresholded. The combination scheme uses the output labels of the classifiers. Soft-level combination, on the other hand, uses estimates of the aposteriori probability of the class. The various combination schemes based on this grouping criterion are shown in Fig. 3.

The sum, product, max, min rules, etc., fall under the soft level combiners as they use the output aposteriori probability of the classifier or a score. Product rule quantifies the likelihood of a hypothesis by combining the aposteriori probabilities generated by the individual classifiers by means of a product rule. Unlike the product rule, the sum rule uses the summation of the aposteriori probabilities. The max rule is an approximation of the sum rule and takes the maximum of the aposteriori probabilities. Similarly, the min rule is an approximation of the product rule. An example of soft-level classifiers combination technique was discussed in [28] for facial recognition. Majority voting is a typical example of hard-level combiners and has found widespread use in the literature. Fig. 4 shows a typical scheme for majority voting.

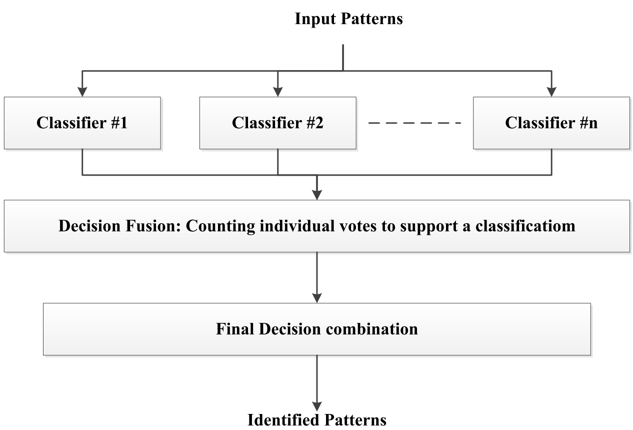

FIGURE 4. Classifiers using majority voting.

There are three different versions of voting: (i) unanimous voting, (ii) more than half voting, and (iii) highest number of votes [8]. Considering the output label vector of the i$^{\text{th}}$ classifier as:

$$[d_{i,1}, \ldots\ldots, d_{i,N}]^T \in [0,1]^N \tag{8}$$

where $i = 1, 2 \ldots, M$ and $d_{i,j} = 1$ if the classifier D_i labels the given pattern as class ω_j, and 0 otherwise. The majority vote results in a decision for class ω_k if:

$$\sum_{i=1}^{M} d_{i,k} = \max_{j=1}^{N} \sum_{i=1}^{M} d_{i,j} \tag{9}$$

Where M is the total number of classifiers, which is usually an odd number, and N is total number of classes. Majority voting provides an accurate class label when at least $^{M}/_{2}+1$ classifiers give correct classifications [29]. The accuracy of the combination scheme is given as:

$$P_{maj} = \sum_{m=\frac{M}{2}+1}^{M} \binom{m}{M} p^m (1-p)^{M-m} \tag{10}$$

where p is the probability of correct classification. Majority voting techniques also require participating classifiers to have comparable accuracies. When the classifiers' accuracies are not similar, it is reasonable to assign more weight to the most accurate classifier. This led to the idea of weighted majority voting. The output label is represented as a degree of support for the different classes.

The discriminant function for class ω_j obtained through weighted voting is:

$$g_j(x) = \sum_{i=1}^{M} b_i d_{i,j} \tag{11}$$

where b_i is the weighting coefficient for classifier D_i. Based on equation (11), the decision rule becomes:

Choose class label ω_k if

$$\sum_{i=1}^{M} b_i d_{i,k} = \max_{j=1}^{N} \sum_{i=1}^{M} b_i d_{i,j} \tag{12}$$

For the sake of convenience, it is a good practice to normalize the weights such that the sum is one.

In weighted majority voting, the weight selection is very important in determining the overall accuracy of the classifier combinations. Therefore, to minimize the classification error of the combination, the weights are assigned as in (13) for M independent classifiers with individual accuracies $p_i, \ldots, p_M$.

$$b_i \propto \log\left(\frac{p_i}{1-p_i}\right) \tag{13}$$

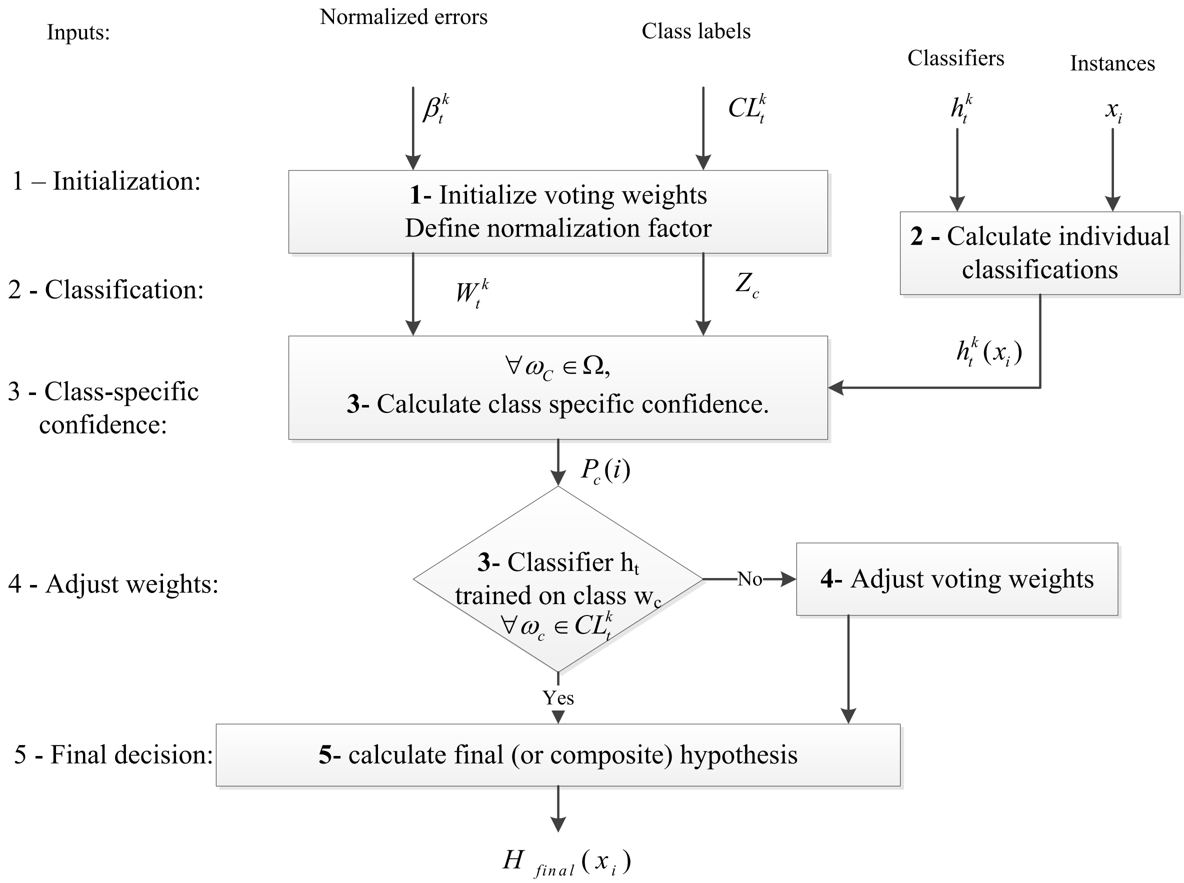

FIGURE 5. Dynamic weight consult-and-vote approach [30].

Besides the weight selection, the prior probabilities of the individual classes are still important in determining the final accuracy. This serves as the background for other modified versions of majority and weighted-majority voting found in the literature. One major advantage of the voting technique is that it allows easy integration of different kinds of classifier architectures.

Muhlbaier *et al.* [30] proposed a method for combining ensembles of classifiers using dynamic weighted consult-and-vote for incremental learning of new classes. The proposed technique is an improvement over a previously developed approach, by the authors, which suffered from the "out-voting" problem in learning new classes. Voting weights are determined by relative performance of each classifier on the training data. The approach learns a new class by allowing individual classifiers to consult with each other to determine their voting weights for each of the test instances. The block diagram in Fig 5, shows the dynamic weighted consult-and-vote approach. The approach shows that classifiers examine each other's decision, cross referencing their decisions with the list of class labels on which they were initially trained, then update the weights.

Another modification of majority voting is to complement it with a divide and conquer technique [31]. Divide and conquer, as the name implies, divides the classification task into a set of smaller and simpler problems, then solving each of problem separately (Fig. 6), followed by a majority voting scheme.

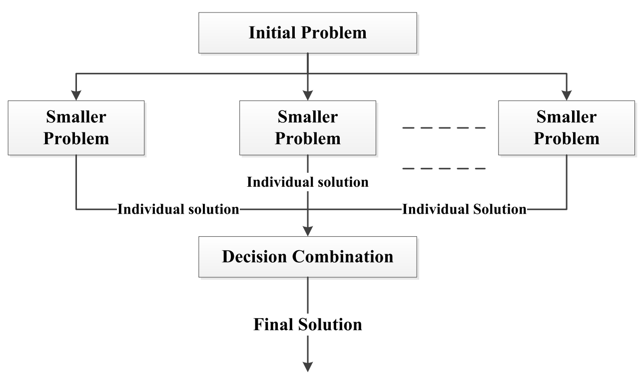

FIGURE 6. Divide and conquer techniques for classifiers combination [31].

Other recent approaches include quality-based combination [32] which gives higher weights to the more reliable classifiers under some given conditions, i.e. better quality. Quality measures are application dependent. In face recognition, for example, common quality measures used include brightness, contrast, and focus, etc.

Simply combining features from different data results in increasing dimensionality, hence escalating the curse of dimensionality. In [33], an effective approach for decision level combination based on spectral reflectance and its higher order derivatives to classify hyperspectral land images was proposed. Their study was carried out under two scenarios: LDA (Linear Discriminant Analysis) based dimensionality reduction followed by a single maximum likelihood classifier, and multiple classifiers decision fusion (MCDF). The techniques were proposed to overcome the curse of dimensionality. Fig 7 shows the block diagram of the MCDF system used in their work.

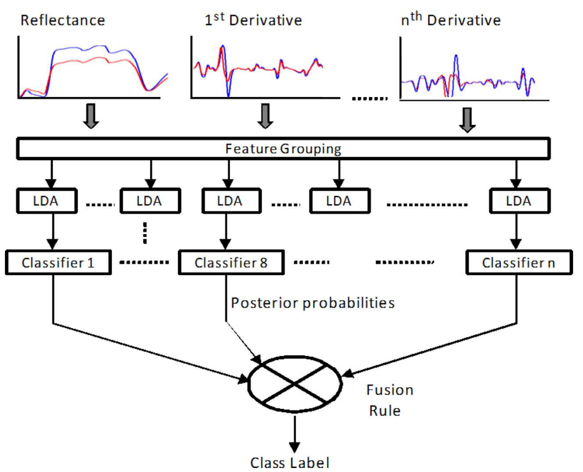

FIGURE 7. **Block diagram of MCDF system [33].**

For the combination of decisions, the authors used a weighted linear opinion pool (WLOP) method, and a weighted majority voting scheme (WMV). The weights were assigned based on the strength of the individual classifiers in identifying the different classes. The authors discussed the benefits of using MCDF over a high-dimensional feature space. Fumera and Roli [7] compared weighted averaging with simple averaging and showed that simple averaging is only optimal when the individual classifiers exhibit comparable performance.

Miller and Yan [34] presented an approach which only tackles two-class problems. The authors introduced a critic associated with each classifier. The purpose of the classifier critic is to predict the error of the classifier. The approach is based on classical standard voting techniques for classifiers combination. It achieves consistent and substantial accuracy improvement over alternative methods on some benchmark datasets. However, it performed best on two-class problem. In [35], an adaptive version of voting techniques was proposed. It involves the weighting of classifiers based on their estimated recognition performance, to generic object recognition problems. The authors applied "Simultaneous Truth and Performance Level Evaluation" (STAPLE). The technique is parameter less, and can be trained with or without labeled data. In [36], three different classifiers (Naive Bayes, J48 Decision tree, and Decision table) were combined using simple, weighted, and probability-based voting. Experiments conducted on the Reuters 21578 datasets showed an improved accuracy when compared with individual classifiers. Weighted and probability voting techniques outperformed simple voting. Additional performance improvement was achieved by applying a supervised dimensionality reduction algorithm.

Magimai-Doss *et al.* [37] proposed a new dynamic entropy-based classifiers combination technique. The technique was tested on the Mandarin TDT4 broadcast news database. The idea of this combination scheme lies in assigning large weights to classifiers that are very confident in their decisions, and small weights to classifiers that are less confident. The approach shows improvement in sentence segmentation when compared with other schemes. In [38], a voting strategy was applied for land cover classification of remotely sensed images using an ensemble of six classifiers. Their results showed an improved performance in image classification. However, this was achieved at the expense of increased computational complexity.

In [39], an approach which combines ANN and KNN based classifiers using majority voting was discussed. This approach was applied to improve accuracy when sensor data is subjected to drift. The dataset used in the experiments contains 13,910 measurements from 16 chemical sensors. Two sets of classifiers were implemented (BP-ANN and KNN). For classifiers combination, a simple median voting was applied for ANN based combination and majority voting for the case of KNN. To test the performance of their setup, data was collected for 37 months, where the data of the first month (batch 1) was used for training and the remaining data was used for testing. Combining several KNN classifiers using majority voting, and median based voting for the ANN classifiers, the authors showed a substantial improvement in performance.

For most multiple classifiers combination techniques, independence between features is usually assumed. Ma *et al.* [40] considered linear dependency of both classifiers and features, in combination techniques. A new framework which models dependency between features without any assumption on feature/classifier distribution was proposed. Two models, Linear Classifier Dependency Modelling (LCDM) and Linear Feature Dependency Modelling (LFDM) were developed. The authors showed that the proposed approach outperformed existing classifier and feature level combination methods under non-Gaussian distribution over four real databases. However, the LFDM technique takes longer time to train when compared to classifier level combination strategies.

In multiple classifiers combination, the individual classifiers either use the same representation of the input pattern or each uses its own representation. Velek *et al.* [41] and de Oliveira, Jr. *et al.* [42] concurred the work of Kittler *et al.* [2] and Kittler [3], showing that the simple sum rule outperforms other rules despite being developed under the most restrictive assumptions. In [43], an approach

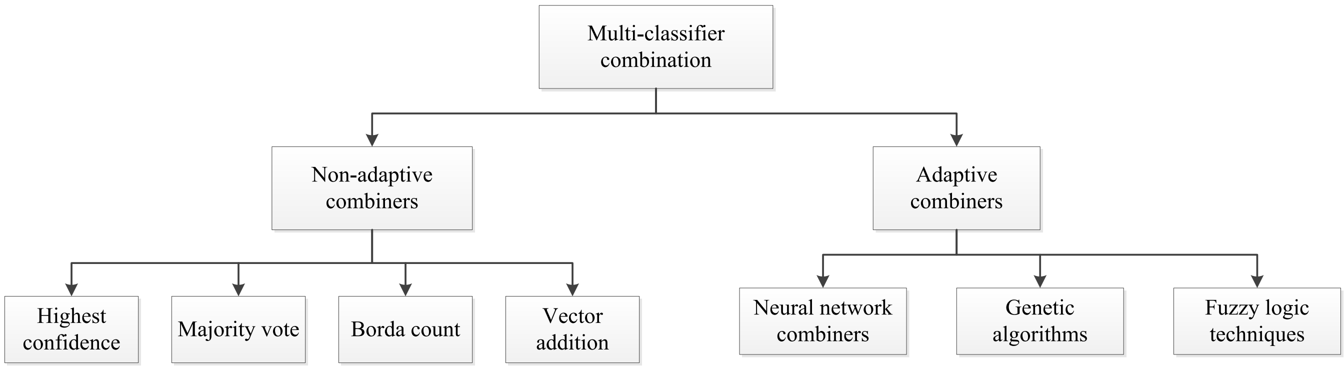

FIGURE 8. **Adaptive and non-adaptive multi-classifiers combination techniques.**

based on majority voting was proposed. Unequal weights are assigned to the base classifiers according to their performance. Such a setup tends to maximize the entropy of probability weights assigned to the individual classifiers subjected to the constrained margin of the ensemble classifier. Experiments were conducted on combining 12 classifiers for digit classification. The classifiers were trained with 50 realizations and tested over 100 test patterns. Voting was found to be the best combination rule for the dataset used.

Fixed rule based combination techniques use class labels, distances, or confidences from the individual classifiers without the need for a training stage. Such rules include product, min, sum, max and median rule. According to [44], fixed rule techniques are only efficient under strict conditions such as availability of large training sets, generation of reliable confidences from base classifiers, and training of the base classifiers on different feature spaces. However, in the absence of the strict conditions, the best result was obtained by a trained combination rule. For example, in the presence of a large training set, the issue of overtraining can be avoided. The base classifiers can be trained on different subsets of data; likewise the combining classifier is trained on a different set.

In [45], a classifiers combination approach was developed for handwritten word recognition. The approach, Runtime Weighted Opinion Pool (RWOP), dynamically assigns weights to the classifiers during runtime and the final combination is weighted according to the local performance of the classifiers. Unlike other weighted sum based approaches, this technique determines the weights with an intuitive run-time strategy. Experimental results performed for recognition of cursive handwritten words demonstrated that the new approach achieves enhanced performance and reduces the relative error rate significantly. Gunter and Bunke [46] proposed a new approach for classifiers combination which strictly work with HMM based-classifiers. Combination of the HMM outputs takes place at a more elementary level unlike some other approaches where the combination is at the decision level. The approach was compared with the max rule and voting strategy. In [12], a new weighted majority voting approach was proposed. The developed approach assigns weights to the different classes rather than the different base classifiers. Such weights were computed by estimating the joint probability distribution of each class with the scores provided by all classifiers in the combining pool. The joint probability distribution was computed using the naïve Bayes probabilistic model. The approach was successfully tested for the recognition of handwritten digits from three standard databases. The major drawback of this approach lies on the statistical independence assumption of variables.

In summary, voting combination strategy and its variants have been widely used in the literature for classifiers combination. Most of the researches have reported good accuracy from the voting technique, an important factor to note is that the base classifiers must be well constituted to achieve a good accuracy.

C. ADAPTIVE AND NON-ADAPTIVE COMBINERS

Adaptive techniques for classifiers combination are mainly based on evolution or artificial intelligence algorithms. They include neural networks combination strategies and genetic algorithms as well as fuzzy set theory. Techniques under these categories are summarized in Fig. 8.

The ANN seeks to simulate how the human brain learns then generalizes the acquired knowledge on unknown test patterns. It has evolved as a very useful classification tool in pattern recognition. It is usually used as a base classifier [29], however, it has also found wide use in combination of classifiers. The ANN consists of many interconnected and identical processing units called neurons [47]. The most widely used ANN model consists of more than one layer and generally is referred to as a Multilayer Perceptron (MLP).

Bogdanov [48] proposed two new approaches for robust and fault-tolerant classifiers combination: Attractor dynamics (AD) algorithm and classifier masking (CM) algorithm. CM is a non-neural version of the AD algorithm based on modeling some properties of sensory integration in the central nervous system. The two approaches are based on the idea of consensus among individual classifiers. The authors show that proposed combination algorithms result in improved robustness and fault tolerance compared with individual classifiers. The developed approaches outperformed the performance of classifiers combination based on averaging and majority voting because the AD and CM algorithms discard corrupted classifier outputs.

Pasquariello *et al.* [49] compared the performance of combining the outputs of an ensemble of ANNs with that of SVM classifiers for remotely sensed data. An MLP module was used for the non-linear combination of the networks outputs. Two approaches were used for coefficient selection optimization: Bayesian and error correlation matrix. Their experimental results showed that the MLP based combination scheme provided the best results.

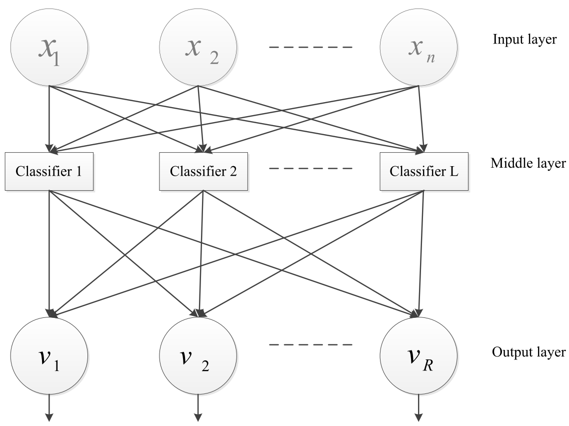

FIGURE 9. **Classifiers combination based on NN.**

Lan and Gao [14] used the ANN, itself, as a model for classifiers combination. In their work, they used different training sets with different classifiers. In addition, the authors integrated the different training sets to train the combination rule. A three-layer ANN (see Fig 9) was used where the different classifiers represent the units of the hidden layer. The approach was tested using the UCI machine learning repository. Results obtained show improved performance over individual classifiers, however, the approach was not compared against other classical classifiers combination techniques.

Di Lecce *et al.* [50] investigated the role of apriori knowledge using existing classifiers combination techniques, namely the Behavior Knowledge space and the D-S theory. The approach performs well when strongly correlated classifiers are combined. Adaptive combiners also include adaptive weighting, associative switching, mixture of local experts (MLE), and hierarchical MLE [51].

In [16], the proposed method combines results from multiple classifiers using a hierarchical architecture, called Fuzzy Stacked Generalization (FSG). To demonstrate FSG, the authors used satellite images segmented and preprocessed to extract different sets of features. Each set is then used with one of K different classifiers. A meta-layer classifier is then constructed by fusing the decisions of base-layer classifiers. Fig 10, shows the architecture of the proposed system. Decisions from base-layer classifiers were aggregated to form a decision vector which is then fed to a meta-layer classifier to make the final decision. The system was tested for detecting buildings from satellite images with improved performance over that of individual classifiers.

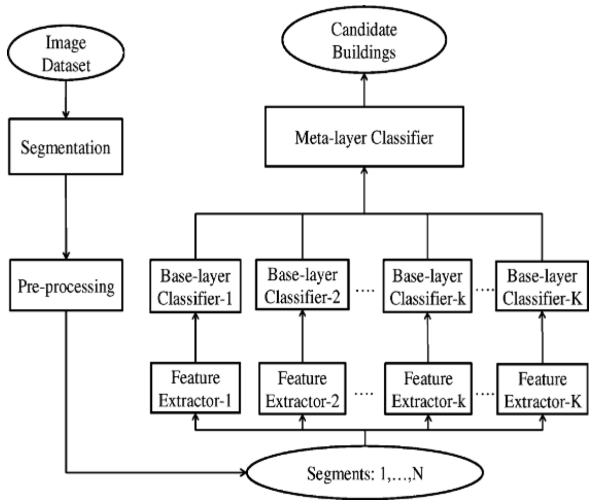

FIGURE 10. **Architecture of decision combination using FSG.**

Zhan *et al.* [22] considered an approach which involves simultaneous extraction and selection of features/classifiers. While GA was used to simultaneously select features/classifiers, the D-S theory was used to combine the outputs of selected classifiers. A typical Korean home environment was used to collect speech signals to evaluate the performance of the proposed system for gender and age classification. Fig. 11, shows the block diagram of the proposed system.

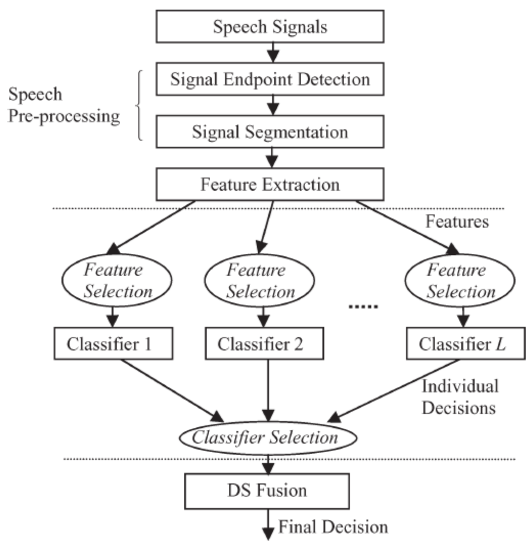

FIGURE 11. **Classifier combination using GA and DS theory [22].**

Classification results obtained from the proposed technique consistently outperformed the individual classifiers results. With D-S fusion, classification accuracy reached 93% when five different classifiers were combined. An improved performance accuracy of 7.5% over the individual classifiers was obtained at the expense of additional computational complexity resulting from the use of the GA.

Kuncheva *et al.* [52] proposed a classifiers combination technique using fuzzy templates (FT). An object is labeled with the class whose fuzzy template is closest to the objects' decision profile. Using two datasets (Satimage and Phonemes) from the ELENA database, the authors obtained an improved performance over majority, min, max and product rules, and unweighted average combination techniques. The authors showed that selecting a subset from the pool of classifiers improves the overall accuracy. In [53], an adaptive fuzzy integral was used to combine multiple classifiers. The parameter λ-fuzzy, which measures performance, is adaptively adjusted depending upon the interaction among the classifiers. The essence of the parameter is to search for the maximum degree of agreement between the conflicting and complementary sources of evidence. On handwriting numeral recognition, their results showed improved performance over conventional combination techniques.

In [54], a fuzzy decision rule was used to combine classifiers. The proposed approach does not require a training stage. In the first step, each classifier is applied separately but no decision is taken. The results from the classifiers are then aggregated using fuzzy decision rule. The class corresponding to the highest membership degree is selected as the correct class. Two measures of accuracy were used: information reliability and global accuracy. These two measures were used in the combination rule, followed by an adaptive fuzzy operator to aggregate the results. An improved accuracy was observed when compared with the individual classifiers. However, the approach was limited to two-class problems. Cococcioni *et al.* [55] proposed a first order Takagi-Sugeno-Kang (TSK) fuzzy model for multiple classifiers combination. The model is an improvement and extension of the linear combination rule. Unlike the classical linear combining methods which assign different weights to each pair of classifiers and classes, their approach associates a weight to each classifier, class, and region of classifier output space (decision boundary). A TSK fuzzy model is further generated to combine outputs of the multiple classifiers. Using the Satimage and Phoneme datasets from the ELENA database, an improved accuracy was obtained as compared to individual classifiers. The approach was also compared with 10 other combination techniques. In most cases, the method is superior to other techniques. However, the authors did not consider possible bias and variance reduction as a result of using the linear model.

In [56], an adaptive approach to classifiers combination was proposed. The approach selects between Bayesian classifiers combination approach and product rule combination strategy. The selection was based on the belief values resulting from the two combination approaches. The approach was compared with a Bayesian method, product, and max rule techniques.

In [57], several variations of the majority voting rule were proposed. A Bayesian framework together with a GA algorithm were used to obtain the weights of the different classifiers. The performance of seven classifiers using these combination schemes over a large set of handwritten numerals was analyzed. An optimal accuracy of 94.3%, 95.4%, and 95.95% were obtained for majority vote, genetic algorithm and Bayesian respectively. In [58], a novel approach based on GA with self-configuration (i.e. the algorithm was designed to re-adjust itself as performance accuracy deteriorates) capabilities was developed. A pool of twelve trained expert classifiers were used to test the approach. The experiments conducted on character recognition (printed and handwritten) showed the benefits of the integration of the GA into the system.

The highest confidence approach is an example of non-adaptive combination techniques. It involves ranking the individual classifiers based on their confidence then selecting the decision of the top ranked one. The use of confidence concept makes the issue of combining classifiers manageable as it allows for continuous feature spaces [44]. An example of confidence based classifiers is a traditional Multilayer Perceptron (MLP) whose confidence value is also obtained with the classification results [51].

The Borda count technique is also an example of non-adaptive methods. It is based on the principle of single winner classifier in which the individual classifiers provide a ranked list of the classes. It is a more sophisticated alternative to majority voting [59] based on ranking level [9]. It does not require training, just like averaging, sum, and voting rules [51]. Plessis *et al.* [60] used the Borda count approach to combine three word recognition techniques. On a lexicon sizes of 10, 100, 1000 words, they achieve accuracies of 88%, 76% and 65%, respectively.

In summary, adaptive combiners tend to do better than the non-adaptive types. This is due to the fact that adaptive combiners update the weights given to the individual classifier dynamically before making the final decision. Given the fact that the performance of the individual classifiers can vary over input patterns, such a dynamic combination provides an edge over its non-adaptive counterpart especially when the data space is wide and diverse.

D. CLASSIFICATION BASED ON THE NUMBER OF CLASSIFIERS

Ensemble-based systems have been introduced recently as a more efficient approach for combining large number of classifiers (usually more than 10) [61], [62]. With these systems, classifiers are either automatically or manually combined into subgroups. An important factor considered in subdividing classifiers into groups is diversity. The most commonly used techniques for ensemble based combinations are displayed in Fig. 12.

Bagging is one of the most intuitive and simple techniques used for ensemble based combination. It is particularly applicable with small size datasets. In Bagging, different datasets are created by bootstrapped versions of the original dataset and combined using a fixed rule like averaging [29], [51].

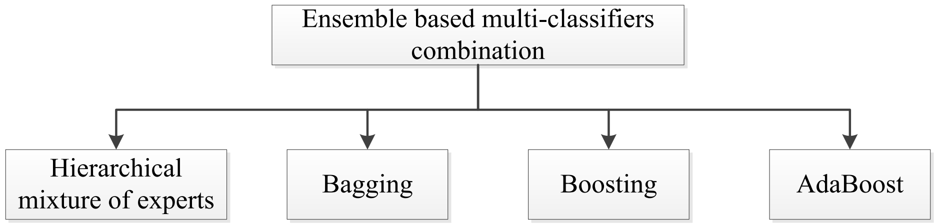

FIGURE 12. **Multi-classifiers techniques using the ensemble-based approach.**

Some variations of bagging include random forest [63], [64] and pasting small votes [8].

However, unlike bagging, in boosting, the individual classifiers are trained hierarchically to discriminate more complex regions in the feature space [9], [51]. These techniques use resampling to create different independent training sets before combination rules are applied. AdaBoost is a variation of the boosting technique. It is an adaptive boosting meta-algorithm that combines outputs of weak classifiers into a weighted sum that represents the final decision. However, the technique is sensitive to noisy data and outliers. Schwenk [65] used AdaBoost to improve the performance of a hybrid HMM/NN speech recognition system. The results showed that AdaBoost performs well even under noisy environments. In [66], a combination approach at features level was considered using SVM classifiers and a Global AdaBoost classifier. Two different training approaches and two datasets were used. The best result was achieved from the combination of features from the two datasets. However, one major drawback of combination at the feature level is the problem of high dimensionality.

E. OTHER COMBINATION TECHNIQUES

In addition to the above approaches, previous literature also includes some special strategies that may not fall under the categories discussed earlier. In [67], for example, a classifiers combination technique based on an SVM active learning algorithm was proposed. Active learning refers to an algorithm that can autonomously select the data points from which it learns [68]. A first level classifier was used to provide class aposteriori probabilities which are then used as inputs to a classifiers-combiner based on SVM active learning. The approach outperforms traditional classifiers combination rules when considering class labeling cost and classification accuracy.

In Ulaş *et al.* [69], [70], the authors proposed the use of eigen-classifiers for combination of correlated classifiers. The proposed technique uses a PCA projection to form uncorrelated eigen-classifiers from a set of correlated classifiers. The idea of uncorrelating classifiers before applying combination techniques makes it possible for the classifiers to complement each other during the combination stage. Thus, the uncorrelation process can be seen as a processing stage before the combination. Results obtained from the proposed approach were compared with Bagging and AdaBoost techniques. The proposed PCA-based technique provided better or comparable accuracy with less number of classifiers as compared to Bagging and AdaBoost. Similarly, in [71], the authors considered the case of linear correlation among the outputs of individual classifiers then uncorrelated them using a simple PCA approach before combination. Following the work in [69], Ekmekci and Cataltepe [71] proposed a generalized kernel based PCA approach to consider non-linear dependencies among the outputs of the individual classifiers. Experiments showed that the generalized kernel based PCA approach outperformed other methods in terms of classification accuracy. In [72], a classifier combination technique based on the extraction of class boundaries and a set of local linear combination rules was proposed. The Phoneme and Ringnorm datasets were used to test the approach. Results obtained were compared with linear combination, voting and decision templates and showed a better accuracy. Moreover, when compared with the KNN rule, the approach was shown to have less computational complexity, however, with a limitation to solving two-class problems only. In [13], a weighted averaging approach was discussed using graph-theoretical clustering and an SVM classifier. The approach was tested on the Oxford flower-17, Event-8, and Scene-15 datasets. Results obtained showed that, the approach, though simple and intuitive, is as powerful as more sophisticated methods. Ho *et al.* [73] used highest rank (HR), Borda count (BC), and logistic regression (LR) for combination of decisions in a multi-classifier system. Decisions of individual classifiers were ranked and the above mentioned techniques were used to either reduce or re-rank a given set of classes during combination. The techniques were used to combine results from four classifiers applied on the degraded machine-printed word recognition problem. Results obtained showed a substantial improvement. A similar approach was also used in [74]. Mixed Group Rank (MGR) was introduced as a new combination technique to balance between preference and confidence by generalizing HR, BC and LR.

Kuncheva [75] discussed an approach which involves switching between classifier combination and classifier selection. The selection was applied on regions over the feature space where one classifier strongly dominates, otherwise the combination is used. The authors further proposed a combination scheme which is a hybrid of clustering-and-selection (CS) and decision template (DT). The approach was compared with: majority voting, Naive Bayes, and simple combination methods such as max, min, averaging.

The authors discussed the tradeoff between selecting the best classifier and combining classifiers. A method also based on classifier selection was proposed in [76]. The authors proposed a method for selection of precise candidate by confidence evaluation of distance-based classifiers. Several rules were proposed for the selection of the precise candidate. Experiments using Euclidean distance and city block for recognition of handwritings showed promising results. Parker [77] used information from the confusion matrix to merge multiple classifiers using class ranking Borda type reconciliation method. Results obtained were compared with three other classifiers combination techniques (majority voting, sum and median rules). Three types of confusion matrices were used: deterministic, uniform, and stochastic. The APBorda (aposteriori Borda count) and sum rule gave the overall best improvement except in the case of stochastic confusion matrix and disparate (classifiers with 10% accuracy difference from each other) combination. An F-measure based combination technique was proposed in [78] with SVM classifier for recognition of human emotion. The F-measure was used in the formation of the decision matrix to determine the final emotion. In [79] the authors proposed an approach combining two different systems for vacant parking space detection. The two systems were fundamentally different in nature as one was based on image data while the second was based on sensor data. The experiments showed that the combination of the two systems provides a reduced error in detecting vacant spaces.

In summary, several classifiers combination techniques have been proposed in the literature with each technique having its own strengths and weaknesses. Recent techniques mostly involve hybridization or modification of previous techniques to achieve better accuracy or to remove an associated constraint on which a particular technique was built on. Some of these constraints include the issue of correlated classifiers, Gaussian distribution, and IID. There is still a need to develop classifiers combination strategies which are not constrained to specific distributions.

IV. DISCUSSION AND CONCLUSION

We presented a comprehensive review of different approaches to classifiers combination. We developed a framework for grouping existing approaches based on level of fusion, type of thresholding, adaptation criteria, number of classifiers, and others. While a large number of combination techniques have been proposed, the literature still lacks a comprehensive performance analysis of such techniques for a given application. The review showed that while one strategy (e.g. fusion at decision level) may outperform others for a given application, the results from such a strategy may not be the best for another application. Therefore, there is a need to try several strategies to find the most appropriate one for that particular application. However overall, it was shown that classifiers combinations in general improve performance significantly over individual classifiers for most problems. Based on our review, we briefly discuss below the main challenges requiring focused research efforts.

An important research direction relies on adding an enhancement stage (post processing) to the classifiers output before applying combination rules. This would improve the performance of individual classifiers before the combination stage. Many classifiers combination techniques have performed well under certain restrictions which include independence assumption, Gaussian distribution, linear process, limited class problem (mostly 2-class problem) and low dimensional feature space. Thus, future work can reconsider relaxing some of these constraints.

The diversity is believed to provide improved accuracy and classifier performance. While this result has been shown by numerous experiments, its theoretical framework has not fully proven yet.

An open area of research involves the use of meta-heuristic algorithms to improve performance of classifiers combination techniques, such as using optimization algorithms with majority voting.

It is important to note that the whole purpose of fusing classifiers is to improve the overall classification performance. The challenge is to show that combining weak classifiers or a bag of both weak and strong classifiers can result in a better performance. Another issue that needs to be further investigated is to explore the advantages of using different strategies for the fusion including probabilistic, learning, decision based, or evidence based techniques.

The discussion on voting based approaches has shown that there is a scope for improving classification accuracy. This issue also offers numerous opportunities for developing optimization techniques to determine the weights. Some of the approaches including GA, PSO, and Ant optimization techniques, among others, can be investigated.

Neural networks as well as similar models such as fuzzy networks, deep neural networks; SVM, etc. also offer an excellent opportunity for developing adaptive techniques to combine individual classifiers outputs. For example, can individual classifiers be considered as layers of more general architectures.

Computational complexity is an important issue that needs further research especially when the different algorithms are deployed over mobile or low power platform.

While in conventional applications the combination of few individual classifiers can enhance the overall performance, the issue of combining a large number of classifiers is still open. Some of the approaches such as hierarchical classifiers combinations decompose the problem into layers of classifiers and the recognition of a given test patterns moves in a top down approach between layers. Another issue of importance is that of combining classifiers when there is a very large number of classes (e.g. more than 500 classes). Deep neural networks have been shown to be able to adapt to such a scenario with a good degree of success.

An important issue which is still open is that of finding the optimal number of classifiers to be combined for a given application. Additionally, for a given number of features, is there a way to distribute these among the different classifiers to be combined.

Finally, the use of hybrid approaches to integrate results from different combination techniques, offers further opportunities for solving more involved applications.

REFERENCES

[1] L. Xu, A. Krzyzak, and C. Y. Suen, "Methods of combining multiple classifiers and their applications to handwriting recognition," *IEEE Trans. Syst., Man and*, vol. 22, no. 3, pp. 418–435, May 1992.

[2] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 3, pp. 226–239, Mar. 1998.

[3] J. Kittler, "Combining classifiers: A theoretical framework," *Pattern Anal. Appl.*, vol. 1, no. 1, pp. 18–27, 1998.

[4] J. Kittler, M. Hatef, and R. P. W. Duins, "Combining classifiers," in *Proc. IEEE ICPR*, 1996, pp. 897–901.

[5] R. P. W. Duin and D. M. J. Tax, "Experiments with classifier combining rules," in *Multiple Classifier Systems*. Berlin, Germany: Springer, 2000, pp. 16–29.

[6] L. I. Kuncheva, "A theoretical study on six classifier fusion strategies," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 2, pp. 281–286, Feb. 2002.

[7] G. Fumera and F. Roli, "A theoretical and experimental analysis of linear combiners for multiple classifier systems," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 6, pp. 942–956, Jun. 2005.

[8] R. Polikar, "Ensemble based systems in decision making," *IEEE Circuits Syst. Mag.*, vol. 6, no. 3, pp. 21–45, Sep. 2006.

[9] S. Tulyakov, S. Jaeger, V. Govindaraju, and D. Doermann, "Review of classifier combination methods," in *Machine Learning in Document Analysis and Recognition*. Berlin, Germany: Springer, 2008, pp. 361–386.

[10] S. Chitroub, "Classifier combination and score level fusion: Concepts and practical aspects," *Int. J. Image Data Fusion*, vol. 1, no. 2, pp. 113–135, Jun. 2010.

[11] H. He and Y. Cao, "SSC: A classifier combination method based on signal strength," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 7, pp. 1100–1117, Jul. 2012.

[12] C. De Stefano, F. Fontanella, and A. S. di Freca, "A novel naive Bayes voting strategy for combining classifiers," in *Proc. Int. Conf. Front. Handwriting Recognit.*, Bari, Italy, Sep. 2012, pp. 467–472.

[13] J. Hou, Z.-S. Feng, and B.-P. Zhang, "A graph-theoretic approach to classifier combination," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2012, pp. 1017–1020.

[14] Y.-D. Lan and L. Gao, "A new model of combining multiple classifiers based on neural network," in *Proc. 4th Int. Conf. Emerg. Intell. Data Web Technol.*, vol. 2. Sep. 2013, pp. 154–159.

[15] H. Kuang, X. Zhang, Y.-J. Li, L. L. H. Chan, and H. Yan, "Nighttime vehicle detection based on bio-inspired image enhancement and weighted score-level feature fusion," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 4, pp. 927–936, Apr. 2017.

[16] C. Senaras, M. Ozay, and F. T. Yarman Vural, "Building detection with decision fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 3, pp. 1295–1304, Jun. 2013.

[17] J. Kittler, "A framework for classifier fusion: Is it still needed?" in *Advances in Pattern Recognition*. Berlin, Germany: Springer, 2000, pp. 45–56.

[18] A. El Khatib, N. Werghi, and H. Al-Ahmad, "Enhancing automatic polyp detection accuracy using fusion techniques," in *Proc. IEEE 59th Int. Midwest Symp. Circuits Syst. (MWSCAS)*, Oct. 2016, pp. 1–4.

[19] M. A. A. Siddiqui, "Fusion of ECG/EEG for improved automatic seizure detection using Dempster Shafer theory of evidence," Ph.D. dissertations, Dept. Elect. Eng., King Fahd Univ. Petroleum Minerals, Dhahran, Saudi Arabia, 2013.

[20] M. Fontani, T. Bianchi, A. De Rosa, A. Piva, and M. Barni, "A framework for decision fusion in image forensics based on Dempster–Shafer theory of evidence," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 4, pp. 593–607, Apr. 2013.

[21] I. Naseem, "Combining classifiers using the Dempster-Shafer theory of evidence," M.Sc. thesis, King Fahd Univ. Petroleum Minerals, Saudi Arabia, 2005.

[22] Y. Zhan, H. Leung, K.-C. Kwak, and H. Yoon, "Automated speaker recognition for home service robots using genetic algorithm and Dempster–Shafer fusion technique," *IEEE Trans. Instrum. Meas.*, vol. 58, no. 9, pp. 3058–3068, Sep. 2009.

[23] S. Saha and S. Saha, "A novel approach to classify edible oil using multiple classifier fusion based on spectral data," in *Proc. Int. Conf. Intell. Control Power Instrum. (ICICPI)*, Oct. 2016, pp. 108–113.

[24] M. Mohandes and M. Deriche, "Arabic sign language recognition by decisions fusion using Dempster–Shafer theory of evidence," in *Proc. IEEE Comput., Commun. Appl. Conf.*, Hong Kong, Apr. 2013, pp. 90–94.

[25] A. Moosavian, M. Khazaee, G. Najafi, M. Kettner, and R. Mamat, "Spark plug fault recognition based on sensor fusion and classifier combination using Dempster–Shafer evidence theory," *Appl. Acoust.*, vol. 93, pp. 120–129, Jun. 2015.

[26] B. Quost, T. Denœux, and M.-H. Masson, "Pairwise classifier combination using belief functions," *Pattern Recognit. Lett.*, vol. 28, no. 5, pp. 644–653, Apr. 2007.

[27] D.-Q. Han, C.-Z. Han, and Y. Yang, "Combination of heterogeneous multiple classifiers based on evidence theory," in *Proc. Int. Conf. Wavelet Anal. Pattern Recognit. (ICWAPR)*, vol. 2, Nov. 2007, pp. 573–578.

[28] R. Toufiq and M. R. Isalm, "Face recognition system using soft-output classifier fusion method," in *Proc. 2nd Int. Conf. Elect., Comput. Telecommun. Eng. (ICECTE)*, Dec. 2016, pp. 1–4.

[29] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*. Hoboken, NJ, USA: Wiley, 2004.

[30] M. D. Muhlbaier, A. Topalis, and R. Polikar, "Learn++. NC: Combining ensemble of classifiers with dynamically weighted consult-and-vote for efficient incremental learning of new classes," *IEEE Trans. Neural Netw.*, vol. 20, no. 1, pp. 152–168, Jan. 2009.

[31] A. Rahman and M. Fairhurst, "Decision combination of multiple classifiers for pattern classification: Hybridisation of majority voting and divide and conquer techniques," in *Proc. 5th IEEE Workshop Appl. Comput. Vis.*, Dec. 2000, pp. 58–63.

[32] N. Poh and J. Kittler, "A unified framework for biometric expert fusion incorporating quality measures," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 1, pp. 3–18, Jan. 2012.

[33] H. R. Kalluri, S. Prasad, and L. M. Bruce, "Decision-level fusion of spectral reflectance and derivative information for robust hyperspectral land cover classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4047–4058, Nov. 2010.

[34] D. J. Miller and L. Yan, "Ensemble classification by critic-driven combining," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 2. Mar. 1999, pp. 1029–1032.

[35] F. Mattern, T. Rohlfing, and J. Denzler, "Adaptive performance-based classifier combination for generic object recognition," in *Proc. Int. Fall Workshop Vis., Modeling Vis. (VMV)*, Nov. 2005, pp. 139–146.

[36] G. Jain, A. Ginwala, and Y. A. Aslandogan, "An approach to text classification using dimensionality reduction and combination of classifiers," in *Proc. IEEE Int. Conf. Inf. Reuse Integr. (IRI)*, Nov. 2004, pp. 564–569.

[37] M. Magimai-Doss, D. Hakkani-Tur, O. Cetin, E. Shriberg, J. Fung, and N. Mirghafori, "Entropy based classifier combination for sentence segmentation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, vol. 4. Apr. 2007, p. IV-189.

[38] L. Dai and C. Liu, "Multiple classifier combination for land cover classification of remote sensing image," in *Proc. 2nd Int. Conf. Inf. Sci. Eng. (ICISE)*, Dec. 2010, pp. 3835–3839.

[39] S. Adhikari and S. Saha, "Multiple classifier combination technique for sensor drift compensation using ANN & KNN," in *Proc. IEEE Int. Adv. Comput. Conf. (IACC)*, Feb. 2014, pp. 1184–1189.

[40] A. J. Ma, P. C. Yuen, and J.-H. Lai, "Linear dependency modeling for classifier fusion and feature combination," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 5, pp. 1135–1148, May 2013.

[41] O. Velek, S. Jaeger, and M. Nakagawa, "A new warping technique for normalizing likelihood of multiple classifiers and its effectiveness in combined on-line/off-line Japanese character recognition," in *Proc. 8th Int. Workshop Front. Handwriting Recognit.*, Feb. 2002, pp. 177–182.

[42] J. J. de Oliveira, Jr., M. N. Kapp, C. O. A. de Freitas, J. M. de Carvalho, and R. Sabourin, "Handwritten recognition with multiple classifiers for restricted lexicon," in *Proc. 17th Brazilian Symp. Comput. Graph. Image Process.*, Oct. 2004, pp. 82–89.

[43] Z. Wu, C.-H. Li, and V. Cheng, "Large margin maximum entropy machines for classifier combination," in *Proc. Int. Conf. Wavelet Anal. Pattern Recognit. (ICWAPR)*, vol. 1. Aug. 2008, pp. 378–383.

[44] R. P. W. Duin, "The combining classifier: To train or not to train?" in *Proc. 16th Int. Conf. Pattern Recognit.*, vol. 2. Aug. 2002, pp. 765–770.

[45] W. Wang, A. Brakensiek, and G. Rigoll, "Combination of multiple classifiers for handwritten word recognition," in *Proc. 8th Int. Workshop Frontiers Handwriting Recognit.*, Aug. 2002, pp. 117–122.
[46] S. Gunter and H. Bunke, "A new combination scheme for HMM-based classifiers and its application to handwriting recognition," in *Proc. 16th Int. Conf. Pattern Recognit.*, vol. 2. Aug. 2002, pp. 332–337.
[47] M. Mohandes, S. Aliyu, and M. Deriche, "Arabic sign language recognition using the leap motion controller," in *Proc. IEEE 23rd Int. Symp. Ind. Electron. (ISIE)*, Jun. 2014, pp. 960–965.
[48] A. V. Bogdanov, "Neuroinspired architecture for robust classifier fusion of multisensor imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1467–1487, May 2008.
[49] G. Pasquariello, N. Ancona, P. Blonda, C. Tarantino, G. Satalino, and A. D'Addabbo, "Neural network ensemble and support vector machine classifiers for the analysis of remotely sensed data: A comparison," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, vol. 1. Jun. 2002, pp. 509–511.
[50] V. Di Lecce, G. Dimauro, A. Guerriero, S. Impedovo, G. Pirlo, and A. Salzo, "Knowledge-based methods for classifier combination: An experimental investigation," in *Proc. Int. Conf. Image Anal. Process.*, Sep. 1999, pp. 562–565.
[51] A. K. Jain, R. P. W. Duin, and J. C. Mao, "Statistical pattern recognition: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 1, pp. 4–37, Jan. 2000.
[52] L. I. Kuncheva, J. C. Bezdek, and M. A. Sutton, "On combining multiple classifiers by fuzzy templates," in *Proc. Conf. North Amer. Fuzzy Inf. Process. Soc. (NAFIPS)*, Aug. 1998, pp. 193–197.
[53] T. D. Pham, "Combination of multiple classifiers using adaptive fuzzy integral," in *Proc. IEEE Int. Conf. Artif. Intell. Syst. (ICAIS)*, Sep. 2002, pp. 50–55.
[54] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "Decision fusion for the classification of urban remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2828–2838, Oct. 2006.
[55] M. Cococcioni, B. Lazzerini, and F. Marcelloni, "A TSK fuzzy model for combining outputs of multiple classifiers," in *Proc. IEEE Annu. Meeting Fuzzy Inf., Process. (NAFIPS)*, vol. 2. Jun. 2004, pp. 871–876.
[56] Y. Yaslan and Z. Cataltepe, "Co-training with adaptive Bayesian classifier combination," in *Proc. 23rd Int. Symp. Comput. Inf. Sci. (ISCIS)*, Oct. 2008, pp. 1–4.
[57] L. Lam and C. Y. Suen, "Optimal combinations of pattern classifiers," *Pattern Recognit. Lett.*, vol. 16, no. 9, pp. 945–954, 1995.
[58] K. Sirlantzis and M. C. Fairhurst, "Optimisation of multiple classifier systems using genetic algorithms," in *Proc. Int. Conf. Image Process.*, vol. 1. Oct. 2001, pp. 1094–1097.
[59] S. Wesolkowski and K. Hassanein, "A comparative study of combination schemes for an ensemble of digit recognition neural networks," in *Proc. IEEE Int. Conf. Syst., Man, Comput. Simulation.*, vol. 4. Oct. 1997, pp. 3534–3539.
[60] B. Plessis *et al.*, "A multi-classifier combination strategy for the recognition of handwritten cursive words," in *Proc. 2nd Int. Conf. Document Anal. Recognit.*, Oct. 1993, pp. 642–645.
[61] E. M. Albornoz and D. H. Milone, "Emotion recognition in never-seen languages using a novel ensemble method with emotion profiles," *IEEE Trans. Affect. Comput.*, vol. 8, no. 1, pp. 43–53, Jan. 2017.
[62] B. Pan, Z. Shi, and X. Xu, "Hierarchical guidance filtering-based ensemble classification for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 4177–4189, Jul. 2017.
[63] L. D. Coronado-De-Alba, A. Rodriguez-Mota, and P. J. Escamilla-Ambrosio, "Feature selection and ensemble of classifiers for Android malware detection," in *Proc. 8th IEEE Latin-Amer. Conf. Commun. (LATINCOM)*, Nov. 2016, pp. 1–6.
[64] M. M. Habib, R. A. Welikala, A. Hoppe, C. G. Owen, A. R. Rudnicka, and S. A. Barman, "Microaneurysm detection in retinal images using an ensemble classifier," in *Proc. 6th Int. Conf. Image Process. Theory, Tools Appl. (IPTA)*, Dec. 2016, pp. 1–6.
[65] H. Schwenk, "Using boosting to improve a hybrid HMM/neural network speech recognizer," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 2. Mar. 1999, pp. 1009–1012.
[66] J. Hu and Y. Chen, "Offline signature verification using real adaboost classifier combination of pseudo-dynamic features," in *Proc. 12th Int. Conf. Document Anal. Recognit. (ICDAR)*, Aug. 2013, pp. 1345–1349.
[67] X. Yi, Z. Kou, and C. Zhang, "Classifier combination based on active learning," in *Proc. 17th Int. Conf. Pattern Recognit. (ICPR)*, vol. 1. Aug. 2004, pp. 184–187.
[68] J. Kremer, K. S. Pedersen, and C. Igel, "Active learning with support vector machines," *Wiley Interdiscipl. Rev. Data Mining Knowl. Discovery*, vol. 4, no. 4, pp. 313–326, Jul. 2014.
[69] A. Ulaş, O. T. Yıldız, and E. Alpaydın, "Eigenclassifiers for combining correlated classifiers," *Inf. Sci.*, vol. 187, pp. 109–120, Mar. 2012.
[70] A. Ulaş, M. Semerci, O. T. Yıldız, and E. Alpaydın, "Incremental construction of classifier and discriminant ensembles," *Inf. Sci.*, vol. 179, no. 9, pp. 1298–1318, 2009.
[71] U. Ekmekci and Z. Cataltepe, "Classifier combination with kernelized eigenclassifiers," in *Proc. 16th Int. Conf. Inf. Fusion (FUSION)*, Jul. 2013, pp. 743–749.
[72] M. Liu, K. Li, and R. Zhao, "A boundary based classifier combination method," in *Proc. Chin. Control Decision Conf. (CCDC)*, Jun. 2009, pp. 3777–3782.
[73] T. K. Ho, J. J. Hull, and S. N. Srihari, "Decision combination in multiple classifier systems," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 1, pp. 66–75, Jan. 1994.
[74] O. Melnik, Y. Vardi, and C.-H. Zhang, "Mixed group ranks: Preference and confidence in classifier combination," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 8, pp. 973–981, Aug. 2004.
[75] L. I. Kuncheva, "Switching between selection and fusion in combining classifiers: An experiment," *IEEE Trans. Syst. Man, Cybern., B, Cybern.*, vol. 32, no. 2, pp. 146–156, Apr. 2002.
[76] C. L. Liu and M. Nakagawa, "Precise candidate selection for large character set recognition by confidence evaluation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 6, pp. 636–641, Jun. 2000.
[77] J. R. Parker, "Combining multiple non-homogeneous classifiers: An empirical approach," in *Proc. IEEE Int. Conf. Cognit. Inform.*, Aug. 2002, pp. 288–292.
[78] A. Agrawal and N. K. Mishra, "Fusion based emotion recognition system," in *Proc. Int. Conf. Comput. Sci. Comput. Intell. (CSCI)*, Dec. 2016, pp. 727–732.
[79] L. Junzhao, M. Mohandes, and M. Deriche, "A multi-classifier image based vacant parking detection system," in *Proc. IEEE Int. Conf. Electron., Circuits, Syst. (ICESC)*, Abu Dhabi, United Arab Emirates, Dec. 2013, pp. 933–936.

MOHAMED MOHANDES received the Ph.D. degree from Purdue University in 1993. He was a Research Fellow with the Center for Sensor Signal and Information Processing, Adelaide, Australia. Since 1997, he has been with the EE Department, King Fahd University of Petroleum and Minerals. His current research interests are in the areas of machine learning, pattern recognition, and Arabic sign language recognition.

MOHAMED DERICHE received the Ph.D. degree from University of Minnesota in 1994. He was with QUT, Australia, where he led the Signal Processing Group. He joined King Fahd University of Petroleum and Minerals, Saudi Arabia. He has published over 200 papers in multimedia signal and image processing. He is the recipient of the IEEE Third Millennium Medal. His research interest is multimedia signal processing with applications in biomedical, seismic analysis, quality assessment, and biometrics.

SALIHU O. ALIYU received the B.Eng. degree (Hons.) in electrical and computer engineering from Federal University of Technology, Minna, Nigeria, in 2010 and the M.Sc. degree in electrical engineering from King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia, in 2015. He is a Lecturer with the Department of Telecommunication Engineering, Federal University of Technology. His research interests include machine learning and multiple classifiers combination for Arabic sign language recognition.