Review on the Architecture, Algorithm and Fusion Strategies in Ensemble Learning

Shruti Asmita
Department of CS
Bansthali University,
Jaipur, Rajasthan, India

K.K. Shukla
Department of CSE
Indian Institute of Technology,
BHU, Varanasi, India

ABSTRACT
Ensemble Learning is an approach in machine learning to find a predictive model taking into considerations the opinions of various experts. Groups of people can often make better decisions than individuals especially when group members come in with their own biases. This document presents a review on the possible architectures that can be used to build an ensemble model, the techniques in which the opinions of the experts could be combined to give a general improved model and the algorithms for implementing the Ensemble Learning. Comparison of architectures is done on the basis of diversity, classification accuracy and memory consumption. This can be helpful in choosing the options depending on the requirement. In the last part an analysis of ensemble learning algorithms on the basis on Bias and Variance is included.

General Terms
Machine Learning, Support Vector Machine, Hard supervised classification.

Keywords
Diversity, Bias, Variance, ensemble learning, classification

1. INTRODUCTION
In the process of making a machine learn something in presence of a supervisor, one train them on several feature inputs providing them with corresponding label. Later one consider the learning good if the classification accuracy of the machine in generating the output as label, taking just feature vector as input, is appreciable. Here this has a single decision maker and it is the single learned entity. Ensemble Learning is the process of training multiple learning machines and combines their outputs, treating them as a "committee" of decision makers. The principle is that the committee decision, with individual predictions combined appropriately, should have better overall accuracy, on average, than any individual committee member [1]. Numerous empirical and theoretical studies have demonstrated that ensemble models very often attain higher accuracy than single models. This ensemble model is also known as the Multi Classifier Systems. For ensemble learning one require some models, a fusion function to fuse all the models, an aggregation method to aggregate the results and total framework architecture. Model is generated by learning functions. The two major measures of analysis of ensemble learning is the classification accuracy and diversity. Classification accuracy is the percentage maximum correct predictions made by the leaned algorithm, on the test data. Diversity is the maximum difference that can be produced in the selection of the subset of the training set of data for training and building each classifier while machine learning. The hypothesis space of an ensemble classifier is wider than the hypothesis space of the single classifier [2]. Some of the classifiers in the ensemble learning can be categorized as Bayes Classifier, Neural Network classifiers, Decision Tree classifiers, K_n Nearest Neighbors classifier, Support Vector Machine (SVM) classifiers Bayes classifier is always an optimal classifier since it has minimum error rate or minimum risk but requires exact knowledge of class prior probabilities and class conditional probabilities of features. It is seldom possible because exact knowledge rarely exists. It has its linear and quadratic variants. The multi-layer perceptron (a non-parametric classifier) is the standard network to use for supervised learning. Training can be very slow, but classification is fast. The number of hidden nodes must be set using validation. Decision tree classifier uses a decision tree and predictive model to output required classifier. When there is large number of features, the decision tree is used to create a model which can be used to classify new vectors accordingly. K_n nearest neighbor classifier is a simple nonparametric method of classification. It has a high computational complexity. The value of k has to be selected by validation. Support vector machine is a hard classifier in which maximize margin approach is used to classify vectors. The aim is to place margin between classes in such a way that the support vectors are placed at maximum distance from the separating margin. Thereby the new test vector classification is done on algebraic comparison with the plot boundary. There are two Key Components of an Ensemble System. First, an algorithm is needed to build an ensemble that is as diverse as possible some of the more popular ones, such as bagging, boosting, Ada Boost, stacked generalization, and mixture of experts [3]. Second, a strategy is needed to combine the outputs of individual classifiers that make up the ensemble in such a way that the correct decisions are amplified, and incorrect ones are cancelled out. Several choices are available for this purpose as well [3].

2. ARCHITECTURE IN ENSEMBLE LEARNING
There are various architectures of ensemble learning depicting the overall framework of the sequence of procedures to be executed to obtain an accurate and diverse ensemble learning system Page Layout.

2.1 Same Classifiers vs. Different Classifiers in Ensemble Learning
The classifier in the ensemble learning has to be combined to get increased classification accuracy. These classifiers can be same for all the sub input set taken or may be different for each input. The feature space for all the input subspace is same.
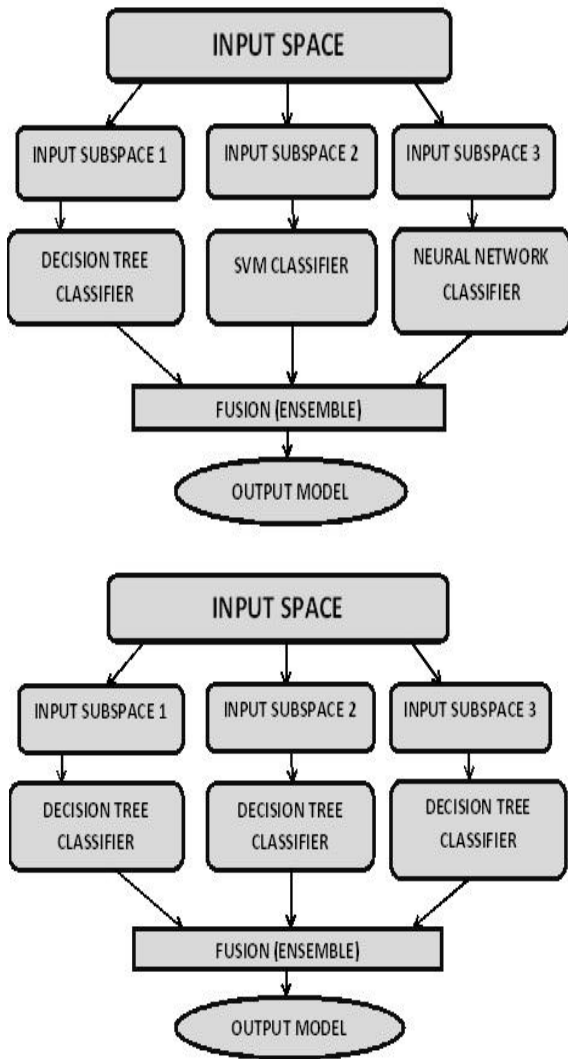
feature fusion is typically to obtain a representation that allows for more effective analysis. On the other hand, the classifier fusion is the method of ensemble of several classifier hypotheses into a single hypothesis by including in it the more effective classifier in greater proportion and the less effective classifier in comparatively less proportion and so on. This results in an overall improved model of classification since the effect of weak classifier has been handled by promoting the effect of strong classifiers. Here the different input subspace fed to the classifiers for model generation before ensemble, have the same feature space. There can also be the scenarios where be require using both the classifier fusion and the feature fusion simultaneously.
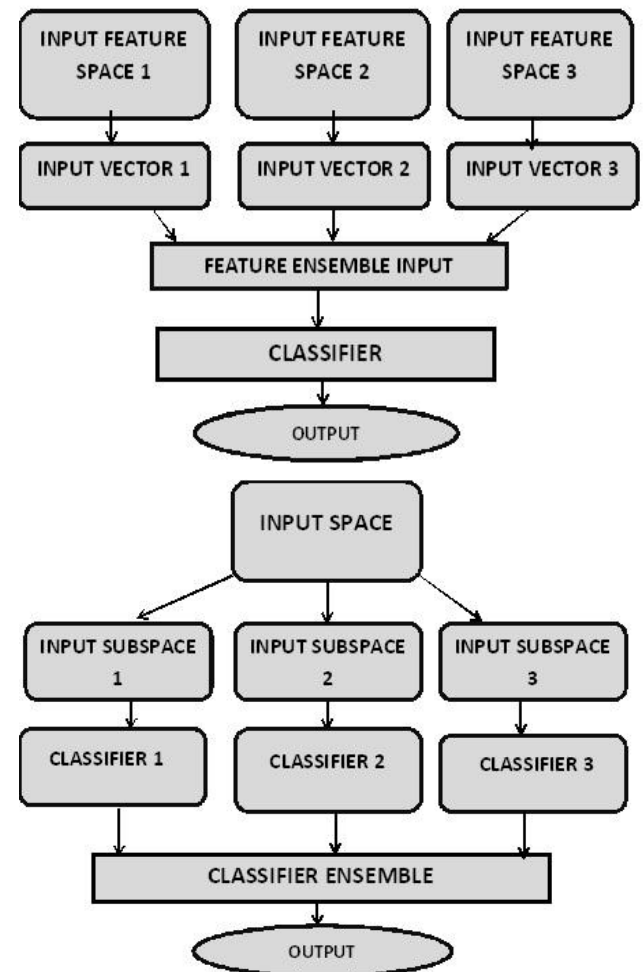

Fig 1: Different classifiers in ensemble (top) and same classifier in ensemble (bottom)

2.2 Feature Ensemble and Classifier Ensemble Architecture

Features are the attribute with corresponding value which describes the input object. For instance, if image classification is considered, pixel's hue, pixel's saturation, pixel's intensity etc. are the features of the image. Features have numerical value associated with them. Feature vector is n dimensional vector of numerical features that represent some object. Feature vector moreover is representation of an object. There can be a scenario in classification where several sets of different features are associated with the object. In such case one requires feature ensemble before predicting the class of the object in classification. Feature ensemble generates and selects single set of features for such set of objects [4]. At the time of ensemble, there may be some features which are not common to all the sets of collected feature vectors. Value of any such useful feature can be calculated by formula utilizing common features over all the sets or can be filled with default values or can be assigned binary values on the basis of threshold over any common feature etc. Any such feature which is not common to majority of feature vectors may be dropped out on the basis of necessity and requirement. Hence in feature fusion, information loss may occur. The purpose of

Fig 2 Feature fusion in ensemble learning (top) and classifier fusion in ensemble learning (bottom)

2.3 Same Feature Input to each Classifier vs Different Feature Input to each Classifier

Each classifier that participates in the ensemble may generate their individual results on the basis of whole feature vector or on the basis of one feature per classifier [5]. The number of classifiers in the second case is many more in number as compared to the first case. The architecture of the first model is shown in Fig 2 (bottom).
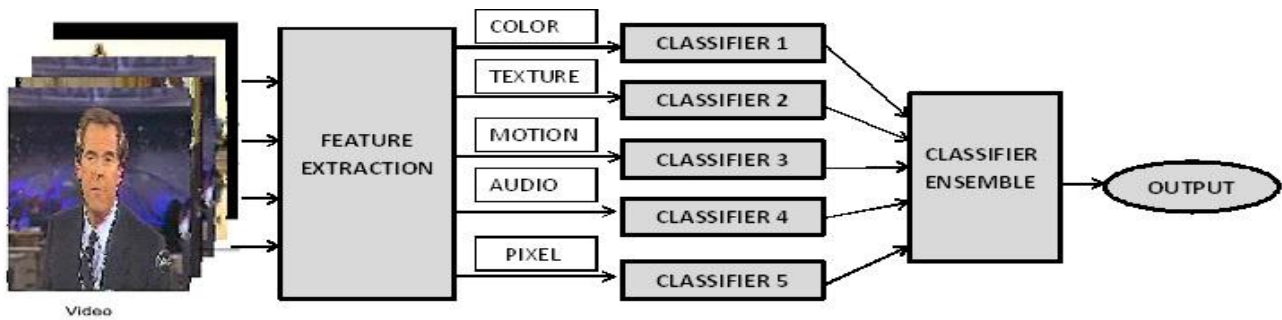
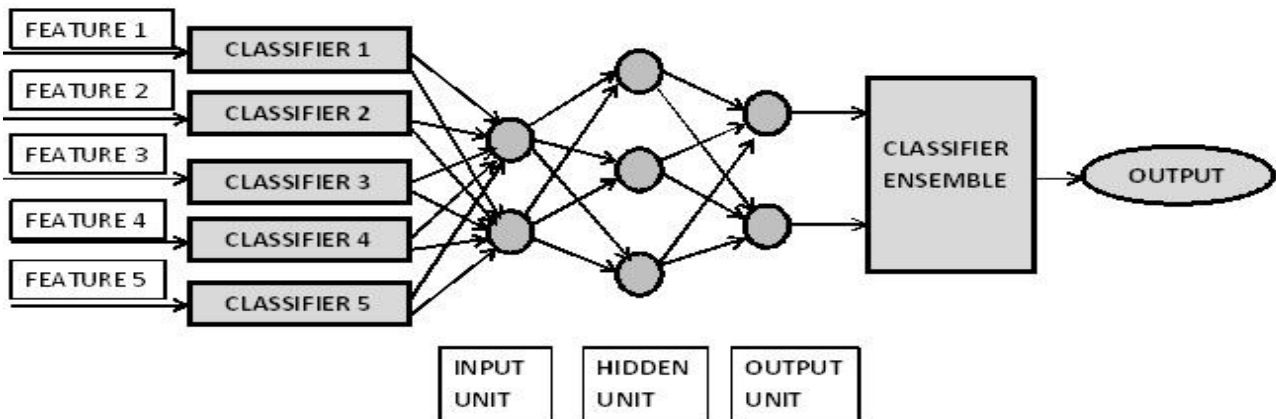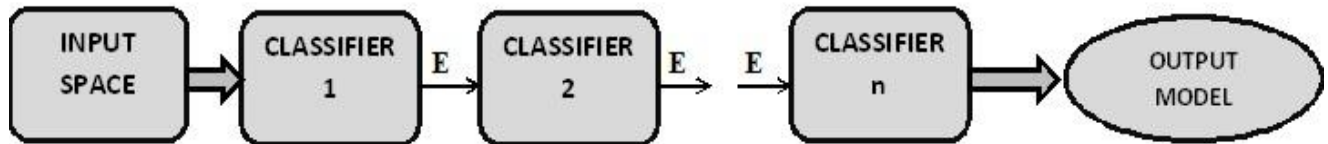**Fig 3 One feature per classifier in ensemble learning**

Fig 4 : Enhance learning by Neural Network usage before classifier ensemble

ENSEMBLE OF ALL PREVIOUS CLASSIFIERS WITH THE PRESENT AT EACH LEVEL (E)

Fig 5 : Sequential fusion of the classifier in ensemble learning

However the architecture of the second model is discussed in Fig 3. The procedure followed in the ensemble of classifiers for both the cases is same. In one feature per classifier architecture (Fig 3), the output of the classifier is fed to the fusion system for a generating ensemble model. A more effective approach can be to first pass the classifier output to the neural network and then the output of neural network can be fed to the ensemble unit. This enhances the learning of overall model produced. The architecture has been discussed in Fig 4.

2.4 Sequential Fusion and Parallel Fusion Architecture

In the parallel fusion topology, the classifiers are fed parallely from the input space and the output model is ensemble of all the classifier models, ensemble parallely at one time. Fig 2 (bottom) shows the parallel fusion architecture. In the serial fusion topology Fig 5, individual classifiers are applied in sequence in some order. When the primary classifier cannot be trusted to classify a given object, e.g., because of the low support/confidence in its result, then the data is feed to a

secondary classifier [6,7,8,9] , and so on, adding classifiers in sequence. The first classifier in the pipeline gives an estimation of the certainty of the classification, so that uncertain data samples are sent to a second classifier specialized in difficult instances. Goal of each classifier is to enhance the accuracy of the previous classifier and balance the lack of previous classifier.

2.5 Online Ensemble Architecture

In all the architectures discussed till now, the data availability is offline i.e. the whole data record is required to be available with the algorithm before start. This demands for a good and large storage unit depending on size and dimensions of data. In real scenarios, there may be a case where wide data is being recorded continuously in each fraction of time. Such data is difficult to store and process. Here comes the concept of online ensemble learning. Online learning algorithms process each training example once "on arrival" without the need for storage and reprocessing, and maintain a current model that reflects all the training examples seen so far [10].

3. ANALYSIS OF ARCHITECTURE ON THE BASIS OF DIVERSITY, CLASSIFICATION ACCURACY AND MEMORY REQUIREMENTS

The primary goal of performing ensemble of classifier is that the effect of classification done by a weak classifier can be balanced by the classification done by a strong classifier. Hence the resulting ensemble model becomes more effective than the individual classifiers. Diversity is a measure of the maximum dissimilarity in the input considered and processed by the individual classifiers for their respective classification so that when the ensemble of these individual classifiers is done the resulting model becomes highly accurate. It is measured as the correlation between two classifier outputs [11,12]. This increases the classification accuracy and decreases the error rate. In the table 1, marked is the diversity, classification accuracy and memory requirement form set {high, medium, low} on the relative basis among the discussed architectures.

4. FUSION STRATEGIES IN ENSEMBLE LEARNING

Fusion unit forms an important component in the ensemble of classifier. The results of the classification produced by each classifier are aggregated at this unit. Each classifier may produce similar or different results for same test data. A good fusion unit must consider the classifications generated by each participating classifiers in the decision making of the ensemble model generation. However it is totally dependent on the algorithm of fusion unit to give those participating classifiers total, partial or no inclusion in the ensemble model, depending on performance of each.

4.1 Evidence based Aggregation

In the ensemble learning, if the information provided by different classifiers is complementary and if numbers of classifiers are combined appropriately, the combined classifier can even outperform the best classifier in the combination [5]. Prediction of the individual performance of the classifier included in the ensemble can result into a good final ensemble model. This prediction can be done by evaluating the evidence value of each classifier (Fig 6). The evidence is the probability of maximum times that a classifier can do correct classification to the total number of classifications required to be done. The evidence value is passed to the decision system. The decision system has right to further allow or stop any classifier based on the evidence and hence the performance, to enter the fusion unit for participating in the contribution towards the ensemble of classifiers [13,14].
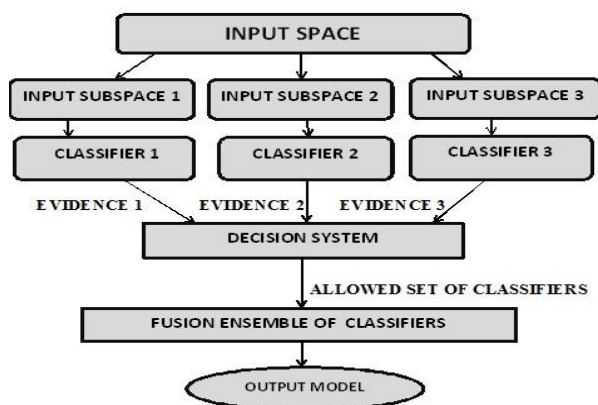

Fig 6. Evidence based architecture in ensemble learning

4.2 Fuzzy Logic based Fusion

Fuzzy logics are used to deal with the imprecision in the world. Basic structure of a fuzzy logic system is composed of four units and they are fuzzifier, fuzzy rule base, fuzzy inference engine and defuzzifier. Fuzzifier maps crisp input values to fuzzy values, fuzzy rule base is a set of IF- THEN rules used for decision evaluation, fuzzy inference engine maps fuzzy input to the rule base to generate fuzzy output values and defuzzifier converts fuzzy output values to a crisp output. In case of hard classification such as support vector machines, there exists a boundary between the classes. In multi class classification, multiple boundaries exist. When ensemble of classifier is taken, different classifier may give different result for same data vector input which can also be complementary to each other. For a particular test data vector, each classifier generates distance from all the boundaries present and the associated classification accuracy This decision is fuzzy (Fig 7) . There are situations like how more or how less classification accuracy is there and similar situation may be how distant or how close the test data is from each boundaries. Suppose there are 3 SVM classifiers each using same or different kernels. There are 3 distances and 3 classification accuracies generated i.e. total 6 input parameters. Distance is having {POSITIVE, NEGATIVE} fuzzy value and classification accuracy is having {HIGH, LOW} fuzzy value associated with them. Total number of IF-THEN rules are 2^6 i.e. 64. The fusion logic system based fusion is shown in fig 9. The defuzzification is done by using formula $y= (\sum_{i=1:64} \beta_i * G_i) / ((\sum_{i=1:64} \beta_i)$ where G_i is the output value of the ith rule and is numerically the value of centroid of the ith isosceles triangle and β_i the firing strength defined by t-norm of membership value of A_j and D_j for j=1,2,3 (in our example case). If this value of y is greater than zero, we consider the data point in positive class else if y is less than zero, we consider the data point in the negative class [11,15].

4.3. Genetic Algorithm based Aggregation System

Genetic algorithm are type of optimization algorithm which use biologically inspired model of intelligence for mimicking the human reasoning[14] The framework of a genetic algorithm aggregation system(Fig 8) shows population which consist of classifiers. The length of chromosome is same as the number of classifiers available for ensemble. Chromosome is a binary string each of whose genes is taken by random generator unit. The random bit 1 corresponds to inclusion of the classifier and random bit 0 corresponds to the exclusion of classifier in genetic algorithm processing. The fitness of the chromosome is decided by the classification accuracy of the classifiers included classifiers. Selection can be done by various available selection algorithms like tournament selection, roulette wheel selection, linear rank selection, and exponential rank selection etc. on the fact of survival of fittest [16]. Crossover is the convergence operator applied to selected parents, to pull the population towards a local minimum/maximum. Mutation is the divergence operator applied after the crossover to occasionally break one or more members of population out of local minimum/maximum space and potentially discover a better minimum/maximum space. In other words crossover facilitates inheritance of characteristics or traits by an offspring from its parents and mutation allows the development of un- inherited characteristics. The chosen crossover probability must be high for fast convergence and chosen mutation probability must be low to provide high diversity. Termination criteria can be

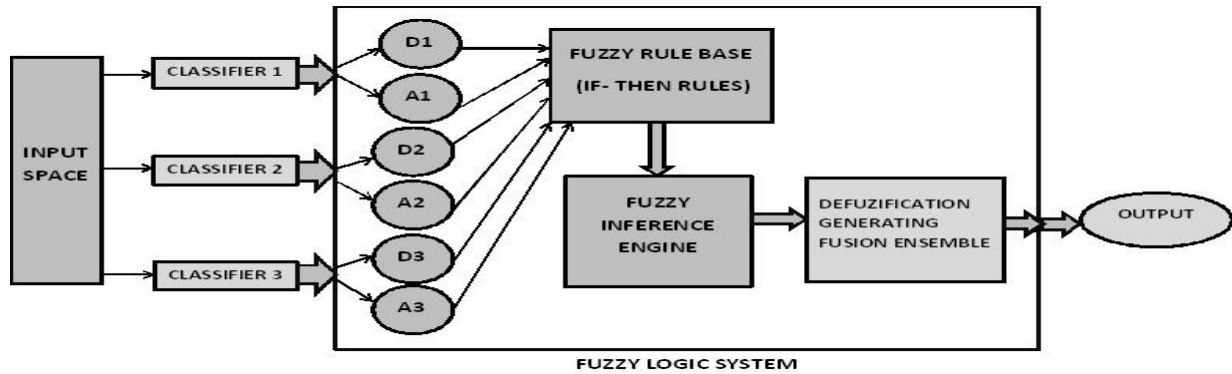maximum number of generations or till the satisfactory solution obtained.

Fig. 7. Fuzzy Logic based fusion unit for generating ensemble of classifier

Table 1 Performance Analysis of architectures in ensemble learning

S.No	Architecture	Diversity in data set (Relative)	Classification Accuracy (Relative)	Space complexity (Relative)
1.	Same classifier ensemble architecture	Medium since the input subspace is selected on random basis.	Medium since the width of classification is restricted.	Medium
2.	Different classifier ensemble architecture or parallel fusion architecture	Medium since the input subspace is selected on the random basis.	High since the width of classification has been increased.	Medium
3.	Feature fusion ensemble architecture	High since the importance of all the features has been worked out in the feature ensemble unit and best is moved forward.	Low since single classifier is used for classification.	High
4.	One feature per classifier ensemble architecture	High since the combination of all the features has been worked out before the classifier ensemble is done.	Low since all the classifiers are not able to classify on the basis of whole set of features	High
5.	Sequential fusion architecture	Low since the single classifier is present at first instant and the random subspace selection for one classifier is less effective than the random subspace selection for a number of classifiers.	High since the next classifier in the series works on ways to better classify the misclassified vectors by previous classifier.	High
6.	Online ensemble architecture	Medium since the data are not stored after learning for any kind of future reference. The set available for the random input subspace selection is restricted to only latest data.	Medium since the ensemble unit cannot make modification to the classifier weights evaluated previously.	Low

4.3. Voting based Aggregation Techniques

4.2.1 Majority Voting

There are three versions of majority voting. Unanimous voting i.e. on which all classifiers agree, simple voting i.e. predicted by at least one more than half the number of classifiers and plurality voting i.e. the highest number of votes, whether or not the sum of those votes exceeds 50% .This is most optimal form of majority voting[15].

4.2.2 Weighted Majority Voting

If we have evidence that certain experts are more qualified than others, weighting the decisions of those qualified experts more heavily may further improve the overall performance than that can be obtained by the plurality voting [15].

4.2.3 Behaviour Knowledge Space based Fusion System

Behavior knowledge space (BKS) method uses lookup table for the fusion of the classifiers. There are a fixed number of classifiers in the ensemble model. Supposing there are 3 classifiers then we have 27 different combinations [17]. During training, we keep track of how often each combination occurs, and the corresponding correct class for each such occurrence. Sample numbers are provided for each combination and for each class. If trained with sufficiently dense training data, and if the table is appropriately normalized, the maximum vote obtained from this BKS table can estimate the class posterior probabilities with reasonable accuracy [15].

4.2.4 *Borda Count*

Borda Count rank the classes in an order of support they receive from the classifiers. If there are N candidates, the first-place candidate receives $N − 1$ votes, the second-place candidate receives $N − 2$, with the candidate in i th place receiving $N − i$ votes. The candidate ranked last receives 0 votes. The votes are added up across all classifiers, and the class with the most votes is chosen as the ensemble decision.

4.2.5 *Dynamic Classifier Selection*

Dynamic Classifier Selection (DCS) methods reflect the tendency to extract a single best classifier from many different participating classifiers [18]. DCS attempts to determine a single classifier, which is the most likely to produce the correct classification label for an input sample [19,20]. Hence only the output of the selected classifier is taken as a final decision.
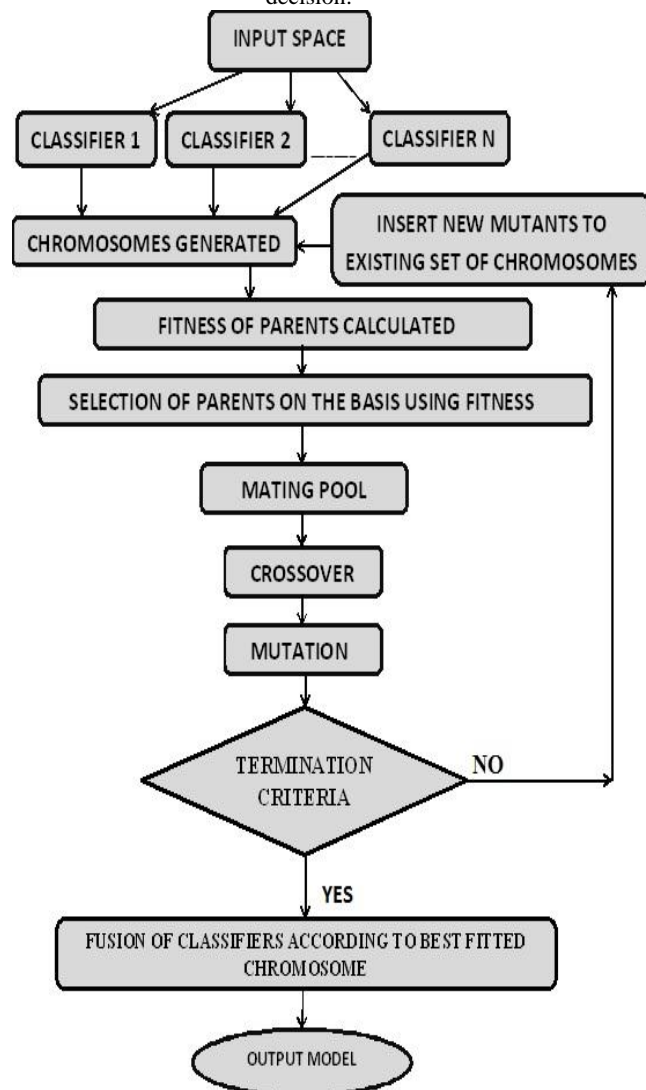

Fig 8. Genetic algorithm based aggregation system in ensemble learning

5. ENSEMBLE LEARNING ALGORITHM

Taking the various architectures, fusion schemes and output combination schemes into consideration we finally have algorithms of ensemble learning which actually work on machine learning platform to give results. Following are popular algorithms in use. Combination of these basic algorithms may be used to generate some new algorithms which may be more effective than the existing. The algorithms in general must reduce the bias and variance to give good classification accuracy.

5.1 Bagging

Bagging is a bootstrap ensemble algorithm that creates individuals for its ensemble by training each classifier on a random redistribution of the training set. The sub input space is selected using technique selection with replacement. Each individual classifier in the ensemble is generated with a different random sampling of the training set. Since bagging resamples the training set with replacement, some instances are represented multiple times while others are left out. Because of this there are chances that for a particular sub input space the classification accuracy may be very high and for other it may be very low. Resampling however is not dependent on the performance of earlier classifiers. Bagging is mostly effective on unstable learning algorithms where small changes in the training set result in large changes in predictions. Neural networks and decision trees classifiers are examples of unstable learning algorithms.

5.2 Boosting

Boosting attempts to produce new classifiers that are better able to predict examples for which the current ensemble's performance is poor [12]. Ada boost is a version of boosting. It generates a set of hypotheses, and combines them through weighted majority voting of the classes predicted by the individual hypotheses. The hypotheses are generated by training a weak classifier, using instances drawn from an iteratively updated distribution of the training data. This distribution update ensures that instances misclassified by the previous classifier are more likely to be included in the training data of the next classifier. [15]. The distribution is initialized to be uniform, so that all instances have equal likelihood to be selected into the first training dataset.

5.3 Stacked Generalization

In a classifier model of classification, there are chances for certain vectors consistently correctly classify since they are away from the decision boundary. Similarly there are chances for certain vectors consistently misclassify since they are close to the decision boundary. For such cases, we can learn certain classifiers that consistently correctly classify or consistently misclassify certain instances by placing two levels of ensemble classifiers.

5.4 Mixture of Experts

In the mixture of expert's algorithm, the experts are trained on different part of the input space and the decision of expert is then combined on the basis of any generalized rule. A gating unit may be placed as a deciding entity for generalized rule application. It works for expectation minimization. Mixture of experts' is a widely used algorithm out of all four mentioned since its easily applicable to wide domains such as bioinformatics, chemical classification, DNA analysis, neurological classification etc.

Table 2 Comment on how each algorithm manages to reduce Bias and Variance to give good classification accuracy.

S. No.	Algorithm	Bias	Variance
1.	Bagging	Bias is reduced because of selection of subsamples with replacement. Hence best set is explored.	Ensemble gives common decision hence the chances of decision varying from each other is reduced.
2.	Boosting	Bias is reduced since all the classifiers work in accordance with the output generated by all previous classifiers.	Variance is reduced since present classifier in its training input set also includes wrongly classified samples of previous classifier.
3.	Stacked Generalization	Placing classifiers at two levels and input of second level is according to output of first level. Hence average of classifier gets more near to target. Bias reduced	The label guess made by classifiers at first level is fed to classifiers at second level. Hence the variance between label outputs of all the classifiers in general is reduced.
4.	Mixture of experts	Gating unit gives a generalization rule to combine the classifiers. Hence average of classifier gets more near to target. Hence bias reduced.	Because of the generalization rule, the difference in guess made by each classifier for the output label is reduced.

6. BIAS AND VARIANCE ESTIMATES

Goal of learning is not only to learn the exact parameters of training and testing data but also to prepare a general model as accurate as possible to predict the unknown labels of any random vector similar to format of learning. A bias term measures difference of the average classifier produced by the learning algorithm from the target function. A variance term measures how much each of the learning algorithm's guesses will vary with respect to each other (how often they disagree). To reduce error we need to reduce both the variance and the bias. Only a probable estimation of bias variance can be done because for actual good values of bias and variance, the training size has to be greatly reduced. Table 2 shows a brief detail on how each ensemble learning algorithm manages to reduce bias and variance terms to give a good classification accuracy.

7. APPLICATION
7.1 Time Series Predictions

Ensemble learning is widely in use for time series predictions in weather forecast, electrical and chemical equipments deterioration compensation, satellite communication, radar and ionosphere data etc. The difference in the time series predictions from the usual predictions lies in the fact that with the time there may occur some drift in the concept or the distribution which is the external factor to compensate in the learning. The best promising way to deal with such issues is found to be the voting fusion strategies i.e. learning latest previous model with larger weights in comparison to the other models in ensemble, can be a solution to this drift compensation.

7.2 Missing Values Prediction

Another significant role of the ensemble learning is in finding the missing data in the real time recordings where uncertainly is also associated. Here one can learn the patterns in the generated readings and accordingly fill the missing values so that the conclusions to be derived from the total data might be accurate.

7.3 Environmental Sciences

Marine sciences are using such machine learning for the algae growth prediction. The feature matrix include the details of temperature level, light exposure etc. Other applications areas in chlorophyll level prediction of leafs, check on the presence of essential or harmful gases , effect of pollutants on the environment belongings etc.

7.4 Media

Media is developing at a great pace. The transformation of the old technologies to the new technologies needs various predictions and learning such as changing the black and white photographs and videos to the color needs learning of the intensity factor of the pixels in the image. This can be well learnt by the ensemble decisions than a single opinion. Various other restoration and preservations also need such learning techniques [22,23].

7.5 Recommendation Systems

Recommendation systems are gaining a markable importance these days. They help consumers to select products, help bidders to invest in stock and shares, advices customers in case of risk associated with switching to the other company for purchases etc. Ensemble learning can provide a diverse opinion in all such cases which outstands the need of recommendation system [24].

7.6 Risk Predictions

Bank credit risk, fraud detection, malware getting corrupt, intrusion detection etc. are such issues which are involved with the finance and hence lie at priority of the customers. Hence to provide better secure environments, the learning is done by taking several opinions instead of one.

7.7 Biometric Predictions

An important application area where the ensemble learning is growing is the biometric recognition especially the handwriting recognition system. Here we can define different classifiers for recognizing the same input and hence consider the opinions of all the decision takers in the final prediction.

7.8. Medical Predictions

Medical field deals with massive data at one or the other instance with very imprecise and ambiguous data at times. In such applications accurate predictions are mandatory. Expert systems such as Computer Aided Diagnosis (CAD) , Clinical Decision Support System (CDSS) and various others that are built for medical diagnoses should also be highly accurate because of the criticality of this application. Hence in this field the system developed which is taking decision through a wide range of opinions is better than a system predicting only on its own individual decision.

8. FUTURE SCOPE

All the above methods are mainly focused on the single level ensemble learning. Double SVM bagging [25] and various such researches show two levels of ensemble of classifiers to achieve even better results in classification accuracy. In the deductions made from the above study and analysis, such two stages algorithms can be developed utilizing same or different fusion techniques in two levels.

Next, in the above discussions the diversity in the data set input to the various learning architectures has been done. However the individual classifiers should also be highly diverse ie the more distant the variety in the approach taken by the classifiers is there, the better the ensemble of classifier results.

Further, the genetic algorithm based fusion strategy can be extended with different combinations of techniques for selection, crossover and mutation operation. The fuzzy logic based fusion strategy has a great further scope in the real time analysis where the natural language input is involved because human being use the vague terms for describing any situation for instance in fuzzy logic the expression cold can be further expanded to very cold, normal cold, severe cold, less cold, more cold etc. to determine the intensity which again vary from individual to individual. Hence such ensemble learning systems which utilize natural language as input are still having a huge scope of development and here we can utilize fuzzy logic based fusion strategy.

8. CONCLUSIONS

In the above content we have discussed various architectures, fusion techniques and various algorithms in ensemble learning. The choice to be made for any scenario is done on the basis of requirement and format of the data available. For instance if feature matrix is a sparse matrix, one can choose classifier per feature architecture. An analysis table1 shows the comparison of diversity and classification accuracy of various architectures on the basis of their respective. Analysis table2 shows how each ensemble learning algorithm manages to decrease bias or variance. The fact that in any scenario one can either have high bias and low variance or low bias and high variance, remains true because of which there lies a bias variance tradeoff in every algorithm. This has been discussed well in the above content. Towards the end, listed are the recent application trends of the ensemble of classifier learning for having better opinions in combined decisions over the individual decisions. The future scope shows the current domains of improving upon the algorithms in this type of learning. Future scope also discusses the critical comments on the existing fusion techniques, architectures and ensemble algorithms and their future possible developments.

9. REFERENCES

[1] Gavin Brown, Encyclopaedia of Machine Learning Chapter No

[2] Pengyi Yang,Yee Hwa Yang, Bing B. Zhou and Albert Y. Zomaya, A review of ensemble methods in bioinformatics.

[3] P´adraig Cunningham , Technical Report UCD-CSI-2007-5. Ensemble Techniques.

[4] Henrik Boström, Feature vs. Classifier Fusion for Predictive Data Mining, a Case Study in Pesticide Classification

[5] De-qiang han, chong-zhao han, yi yang, Multi-class svm classifiers fusion based on evidence combination

[6] Rachid Benmokhtar and Benoit Huet, Classifier Fusion: Combination Methods For Semantic Indexing in Video Content

[7] L. Lam, Classifier combinations: implementations and theoretical issues, in: Proceedings of the First International Workshop on Multiple Classifier Systems

[8] A.F.R. Rahman, M.C. Fairhurst, Serial combination of multiple experts: a unified evaluation, Pattern Analysis and Applications.

[9] Michał Wozniak, Manuel Grana, Emilio Corchado, A survey of multiple classifier systems as hybrid systems.

[10] G.T. Prasanna Kumari, A Study Of Bagging And Boosting Approaches To Develop Meta-Classifier

[11] De-qiang han, chong-zhao han, yi yang, Multi-class svm classifiers fusion based on evidence combination .

[12] Khalid Jebari, Mohammed Madiafi , Selection Methods for Genetic Algorithms

[13] G.T. Prasanna Kumari, A Study Of Bagging And Boosting Approaches To Develop Meta-Classifier

[14] Tom Dietterich, Rich Maclin, Bias Variance Tradeoff and Ensemble methods

[15] Xiujuan Chen, Computational Intelligence Based Classifier Fusion Models for Biomedical Classification Applications

[16] Muhammad Nazir, Arfan Zaffar, Ayaz Hussain, M.Mirza, Efficient gender classification using optimization of hybrid classifiers using genetic algorithm

[17] Robi Polaker, Ensemble based systems in decision making.

[18] H. Ichihashi, T. Shirai, K. Nagasaka, and T. Miyoshi, Neuro-fuzzy ID3: A method of inducing fuzzy decision trees with linear programming for maximizing entropy and an algebraic method for incremental learning.

[19] Dymitr Ruta and Bogdan Gabry, An Overview of Classifier Fusion Methods.

[20] T.H. Ho, J. J. Hull, S.N. Srihari, Decision Combination in Multiple Classifier System.

[21] K. Woods, W.P. Kegelmeyer, K. Bowyer, Combination of Multiple Classifiers Using Local Accuracy Estimates

[22] Li Tan, Yuanda Cao, A novel fusion method for semantic concept classification in video.

[23] Rachid Benmokhtar , Benoit Huet, Perplexity based evidential neural network classifier fusion using MPEG-7 low level visual features.

[24] Michal Wozniak, Manuel Grana, Emilio Corchado, A survey of multiple classifier systems as hybrid systems.

[25] Faizal M. Zaman, Hideo Hirose, Double SVM bagging, a new bagging with support vector machines.