Intelligent Object and Pattern Recognition using Ensembles in Back Propagation Neural Network

Hafiz T. Hassan, Muhammad U. Khalid and Kashif Imran

Abstract— Object and Pattern Recognition applications are one of the most active research fields in the Neural Network research community today. Neural networks are mostly preferred in predictive and classification problems because they are unstable classifier that is little change in input causes drastic changes in output. Now in some scenarios the Neural Networks are not enough to do the required job such as classification or regression analysis so an evolved architecture called ensembles is used that is instead of one classifier such as a neural network many are used. In the paper we are going to incorporate Back Propagation Neural Network for the classification of two datasets (the sonar and ionosphere datasets) in bagging ensemble architecture using PCA and standalone architecture.

Index Term— Back Propagation, Bagging, Classification, Ensemble, Neural Network.

I. INTRODUCTION
The recognition field has become so vast that it has been divided in numerous small fields two of the most important are emphasized in this paper and defined as follows. Pattern Recognition is the science of classification of different entities. Object recognition is the task of finding an object present in some data. The data can be a numeric matrix or vector based data or it could be an image [1].

Conventional techniques used for recognition have become obsolete so now Artificial Intelligent systems are used. Artificial intelligence [2] during the last two decades has opened new prospects for the modern research era. Numerous fields like signal processing, image processing, bioinformatics, power, space exploration, weather predictions and financial forecasting etc are using Artificial Intelligence based systems. Where we have too little or too much data or the data type is such for instance in complex terms we require intelligent systems. The main reason being humans no matter how much IQ we have, we cannot process a large amount of data or make conclusions based on that data and cannot focus for a long time.

Artificial intelligent systems are also termed as classifiers. The most notorious and widely used classifiers are neural networks [3], decision trees [3], [2], genetic algorithms, expert systems [3] etc.

Now for this paper we focus solely on neural networks. “A Neural Network is basically a statistical computational model of the human biological nervous system”. The researches of using neural networks in object and pattern recognition problems dates back almost to the formation of first neural network.

Now, the question arises why we are using neural networks? The answer is quite simple because of their flexibility and their tendency to predict, classify all sorts of data better than any other classifier. Neural networks are basically classified in to two categories,

1- Feed-Forward Neural Networks
2- Recurrent (or Feedback) Neural Network

Our datasets classification problem exhibit outputs at two levels; such problems are termed as binary classification problems. And for cases the feed-forward neural networks are more viable. Now in feed-forward networks we have used the Feed-forward Back propagation Neural Network [2]. The two datasets we are using for the paper are the sonar dataset and the ionosphere dataset taken from the UCI repository of Machine Learning [4].

There are many scenarios as discussed below due to which the decision of one neural network is incapable of giving a confident decision.

There are certain classification problems and scenarios, in which single classifiers don’t perform very well like,

• Presence of too much data.
• Inadequate availability of data.
• Data having complex class separations.
• Data Fusion.
• Generalization Problem.
• Insufficient accuracy for some delicate fields such as medical diagnosis and defense scenarios.

So to overcome such an incident ensembles are used in integration with a classifier (like Neural Network) and Machine Learning techniques.

This paper is organized as follows. Section 2 briefly describes the Feed-forward Neural Network. The datasets are described in Section 3. Section 4 gives the description of ensembles. Finally, Section 5 gives Simulations and results. Section 6 is comparative analysis and 7 accounts for conclusion.

II. FEED-FORWARD BACK PROPAGATION NEURAL NETWORK
Back propagation Neural Networks is a network that based on Back propagation learning technique and that works using supervised learning [3]. In general it is called the Feed Forward Back propagation neural network. With regard to architecture it is basically a Multi-layer Perceptions [2]. The
Back propagation neural network was the gemstone that enchanted and mesmerized researchers and showed the true power of neural networks. It opened research doors with endless opportunities in various fields of engineering, sciences and statistics; and it is computationally economical. But on the darker side the BPNN has also been called the ‘black box’ as it has a fixed algorithmic operation and that’s about it, there is no fixed topology (number of nodes and neurons used) for it, it exhibits different results with different data subsets within the same dataset that is it is very hard to sway it to global minima. Regardless of all these factors overall the BPNN is quite accurate and easy to manipulate with respect to other neural networks.

A. Steps involved in Working of BPNN
- The dataset instances are mixed to provide most generalized results.
- The dataset is divided into two parts; the training and the testing datasets.
- The training dataset comprises of about 70% of the original dataset and the testing dataset occupies about 30%.
- Now the neural network is trained on known target value (in our case binary).

After the BPNN has been trained, the testing dataset that the neural network has never seen is applied to check the accuracy of the classification. The fig 1 shows a basic BPNN comprising of an input, hidden and output layer.

![Simple Back propagation Neural Network](image)

Where a1, a2 and an are the inputs applied on the input layer, n1 is the hidden layer and n2 is the output layer. The output of the network is ‘y’. ‘δ out’ is the error signal that is generated when the output ‘y’ is compared to the target output of the training dataset comprising of the ideal classification result. The error signal moves from the output layer to the hidden layer changing the weights to adjust to the correct result once this error is minimized close to zero the weights are fixed meaning the network is trained and can be tested.

III. DESCRIPTION OF DATA SETS
A. The SONAR Dataset
SONAR stands for ‘Sound Navigation and Ranging’ and is used by submarines or other vessels mainly to detect underwater mines, other submarines and destroyer ships.

![On board system must be able to detect the mine correctly.](image)

We have taken the sonar data from UCI repository of Machine Learning [4]. The sonar dataset comprises of 208 vectors having 60 attributes or it can be said that the data is 60 dimensional where each value signifies the strength of the signal at a particular frequency over a span of time; these frequencies increase in one vector meaning the last value is taken at highest frequency. The first 97 vectors represent rocks at different aspect angle attained by bouncing sonar signals from rough rock surface. Next 111 vectors are Mines attained at different aspect angles by bouncing sonar signals over a metallic cylinder. In the target of the training data mines are represented by 0’s and rocks by 1’s.

For the sonar dataset the distribution of data used is shown in Table I.

<table>
<thead>
<tr>
<th>Overall Data (208)</th>
<th>Training Data (146)</th>
<th>Testing Data (62)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rocks</td>
<td>Mines</td>
<td>Rocks</td>
</tr>
<tr>
<td>97</td>
<td>111</td>
<td>66</td>
</tr>
<tr>
<td>30</td>
<td>32</td>
<td></td>
</tr>
</tbody>
</table>

Data distribution is not fixed, that is why different researchers get different results and thus accuracy varies.

Now to get a rough account about the complexity of the class boundary between the rocks and mines, one instance of each is plotted as an image normalized to fit the entire color range. Rock instance is plotted in Fig. 3.
Similarly the mine instance is plotted in Fig. 4.

Fig. 4. Graphical Image representation of mine instance.

Now both Fig. 3 and Fig. 4 clearly show that there is very little difference in the band colors with respect to the band colors and dimensions. So it clearly shows that the class boundary is very close between the two entities.

B. The IONOSPHERE Dataset

The second dataset used is the ionosphere dataset and it also has been taken from the UCI repository [4].

The dataset comprises of 351 instances having 34 attributes each, where every instance shows radar returns classified as good and bad. Good radar returns show the presence of free electrons and bad show there absence. In other words the main objective is to predict at which instances the ionosphere is structured using the BPNN. In the target sets good returns are denoted by ‘0’ and bad returns by ‘1’. The training dataset comprises of 275 instances and the testing dataset comprises of 76 instances. Here the defense aspect is a little shallow but it does the job.

For the ionosphere dataset the distribution of data used is shown in Table II.

<table>
<thead>
<tr>
<th>Overall Data (351)</th>
<th>Training Data (200)</th>
<th>Testing Data (151)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bad</td>
<td>Good</td>
<td>Bad</td>
</tr>
<tr>
<td>126</td>
<td>225</td>
<td>99</td>
</tr>
<tr>
<td>Good</td>
<td>101</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td></td>
<td>124</td>
</tr>
</tbody>
</table>

As was done for the sonar the ionosphere instances are also plotted as images as shown in Figures 5 and 6.

Fig. 5. Image of good radar return instance

Fig. 6. Image of bad radar return instance

Now in this case both instances are very different from each other and thus even by looking at the images they can be distinguished.

IV. ENSEMBLES

A. Introduction to Ensembles

In subjects of great significance that posses medical, pecuniary or public values, it is useful to take different opinions or views before making a decision. Considering the case of a patient suffering from any disease, for treatment the doctor suggests surgery that is risky, the patient will take second, third multiple opinions and weigh them according to the different doctor’s experience and qualification before coming to a final decision. That is human nature consulting with people that are experts in the relevant areas to get some sort of assurance and mental satisfaction in making a decision. So it is no surprise that same procedures and techniques are being adopted in automatic decision making systems [5].

An ensemble consists of a set of independently trained classifiers whose predictions are combined by various statistical or algebraic methods [6]. The classifiers that incorporate ensembles are fuzzy logic, decision trees, neural networks, expert systems etc. For this thesis neural networks are used [6]. Ensemble learning is based on manipulation of three entities; the input data, the model, the output. In general a neural network ensemble is constructed in two steps,

1- Training a number of component classifiers like neural networks
2- Combining the component predictions

Re-sampling is an input manipulation technique that can
be achieved by a variety of methods. The re-sampling methods in ensembles are widely adopted as they are flexible in handling as the architecture or we can say topology of the classifiers in not changed only the input data is manipulated. The most popular and widely used re-sampling technique is bagging [8].

B. Why use Ensembles?

There are numerous reasons for using ensembles some are discussed below,

i) Statistical Reasons
When dealing with classifiers such as Neural Networks it is evident that good training performance of a particular dataset does not ensure good generalization [6]. This is due to many reasons like the training data does not fully represent the whole data. In such cases thus combination of neural networks can be used that can decrease the risk of bad generalization.

ii) Inadequate availability of Data
Some datasets are too small and thus if classifiers are used there is inadequate learning and thus under-fitting occurs. Ensemble techniques involving data re-sampling are very helpful in the case, each re-sampled set is trained on an individual classifier combined to get final result.

iii) Presence of too much Data
In some problem the amount of data to be handled is too large so it is not feasible to use a single classifier because it will be a very time consuming task processing bulky data also special computer will be required for the task. Data in some cases can range up to tens of Gigabytes. So this problem is approached by dividing the data in to smaller datasets that are trained on individual classifiers and the results are then combined.

iv) Data Fusion
In many cases additional information is available that cannot be given to the same classifier tackling the core information so the additional information is trained separately and then combined [5].

C. Applications of Ensembles

Ensembles have applications in all fields where classifiers are used and where some problems discussed earlier exist. Some field that use ensembles are,

- Medical diagnosis
- Image and Video Recognition
- Signal Processing
- Inspection of Oil/Gas lines
- Financial forecasting

D. Diversity among Ensembles

Ensembles are used because no single classifier with perfect generalization performance is present. Factors like noise, complexity are a reality so researchers resort to ensembles [9], [10].

So the goal must be to make the individual classifiers as diverse as possible because if they are alike than what is the point of ensembles, they will reach the same conclusion that could be reached by a single classifier. One widely used effective method of making different datasets out of the original dataset and then trained on individual classifiers. The most popular re-sampling technique for making multiple datasets is called bagging [11].

E. Bagging

Bagging, which stands for ”bootstrap aggregation”, is an ensemble method of combining multiple predictors [8]. The bootstrap is a procedure that involves choosing random samples with replacement from a data set [7]. Sampling with replacement means that every sample is returned to the data set after sampling. So a particular data point from the original data set could appear multiple times in a given bootstrap sample. The resultant bootstrap has the same dimensionality as the input data. Bagging is very effective for classification problems thus used for both datasets. Now other than these machine learning tools PCA [12], [14] is also used for dimension reduction [13] to give another frontier for comparative analysis.

V. SIMULATIONS AND RESULTS

For the validity of results and comparisons, simulations were performed using different scenarios. The main goal of this thesis was to get classification using ensemble technique called bagging using neural networks for two datasets. The following simulations were performed,

1. Classification of Original datasets using BPNN
2. Classification of 3-Bagged ensemble datasets using BPNN
3. Classification of 25-Bagged ensemble datasets using BPNN
4. Classification of 25-Bagged ensemble dataset with PCA using BPNN

For all simulations Matlab 7 was used. As mentioned before the neural network used is the Back propagation Neural Network (BPNN) for all scenarios.

A. Classification of Original Datasets using BPNN

The sonar dataset is first trained according to distribution showed in the dataset section and after training on the BPNN the testing data is applied to the trained neural network the output is compared to actual output and error is calculated that is plotted in Fig. 7 and accuracy of classification is given in Table III.

![Fig. 7. Classification result of sonar dataset](image-url)
Table III
Accuracy of Classification for Sonar Dataset

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy of mine</td>
<td>90.62%</td>
</tr>
<tr>
<td>classification</td>
<td></td>
</tr>
<tr>
<td>Accuracy of rock</td>
<td>86.67%</td>
</tr>
<tr>
<td>classification</td>
<td></td>
</tr>
<tr>
<td>Overall dataset accuracy</td>
<td>88.709%</td>
</tr>
</tbody>
</table>

The same procedure is applied for the ionosphere dataset and the output is shown in Fig. 8.

Fig. 8. Classification result of ionosphere dataset

Table IV
Accuracy of Classification for Ionosphere Dataset

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy of bad return</td>
<td>77%</td>
</tr>
<tr>
<td>classification</td>
<td></td>
</tr>
<tr>
<td>Accuracy of good return</td>
<td>95%</td>
</tr>
<tr>
<td>classification</td>
<td></td>
</tr>
<tr>
<td>Overall dataset accuracy</td>
<td>92.058%</td>
</tr>
</tbody>
</table>

B. Classification of 3-Bagged ensemble datasets using BPNN

This section is the start of incorporating ensembles modifying the previous model by using bagging technique. Details of the technique are discussed in the ensemble chapter. To start things of first a smallest ensemble model (that is three) is constructed. As the datasets are binary classification sets the number of bagged replicates taken is odd otherwise the result could en up to be fifty-fifty. And the first odd number greater than one is three.

The block diagram of 3 bagged ensemble BPNN is shown in Fig. 9.

Fig. 9. Block diagram of 3 bagged ensemble BPNN

Now for both datasets bootstrap replicates are formed using 90% confidence estimation.

First the sonar dataset is used in accordance to the steps shown in the above block diagram. The final result is shown in Fig. 10 and accuracy of classification is given in Table V.

Fig. 10. Classification result of 3 Bagged Sonar dataset

Table V
Accuracy of Classification for Sonar Dataset

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy of mine</td>
<td>80%</td>
</tr>
<tr>
<td>classification</td>
<td></td>
</tr>
<tr>
<td>Accuracy of rock</td>
<td>81.25%</td>
</tr>
<tr>
<td>classification</td>
<td></td>
</tr>
<tr>
<td>Overall dataset accuracy</td>
<td>80.645%</td>
</tr>
</tbody>
</table>

Similarly for the ionosphere dataset final result is shown in Fig. 11 and accuracy of classification is given in Table VI.
Now in this case the accuracy instead of increasing has decreased because the bootstrap combinations were not sufficient enough for adequate learning and thus giving high accuracy results.

C. Classification of 25-Bagged ensemble datasets using BPNN

The next model uses 25 bootstrapped replicates therefore 25 BPNN’s will be used. This is a long and hectic process as one misstep in the process and the overall result will change. Again just like the case of 3-bagged ensembles here also the testing data remains untouched and the training data is replicated 25 times. The number of bootstrap must again be odd for one entity for both datasets must dominate in terms of votes to get the final classification that in case of even number can result in a tie. In case of using even numerous other algorithms are used by researchers such as Genetic Algorithms but that makes the solution very complex. In ensembles in such cases one more thing can be done that is to give an account for probability for every ensemble component neural networks for that a lot of knowledge about the data must be available. It is again a lengthy and complex process and researchers avoid it mostly.

The block diagram of 25 Bagged BPNN is shown in Fig. 12.

The result obtained for 25-Bagged Bootstrapped BPNN for sonar is shown in Fig. 13 and accuracy of classification is given in Table VII.

![Fig. 12. The block diagram of 25 Bagged BPNN.](image)

Table VI

<table>
<thead>
<tr>
<th>Accuracy of Classification for Ionosphere Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy of bad return classification</td>
</tr>
<tr>
<td>Accuracy of good return classification</td>
</tr>
<tr>
<td>Overall dataset accuracy</td>
</tr>
</tbody>
</table>

Fig. 13. Classification result of 25 Bagged sonar dataset

![Fig. 13. Classification result of 25 Bagged sonar dataset](image)

Table VII

<table>
<thead>
<tr>
<th>Accuracy of Classification for Sonar Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy of mine classification</td>
</tr>
<tr>
<td>Accuracy of rock classification</td>
</tr>
<tr>
<td>Overall dataset accuracy</td>
</tr>
</tbody>
</table>

The majority voting decrease the variance of the data thus the bias is increased and the overall system gets closer to optimal fitting. The above statement holds true only if the variance is decreased to some extents otherwise the bias will increase too much and under-fitting will occur and vice versa over-fitting will occur.

Also the ensemble parameters such as number of bootstraps are not fixed and there is no concrete way or formulae that can find the best number, the response must be plotted and explored if possible to get an estimate. Secondly in bagging the attributes of data instances are chosen randomly with each attribute has the same probability to be chosen so the good performance of bagging still remains a mystery, the best possible explanation being the bias-variance tradeoff.
Now moving to ionosphere the same model is applied and result is shown in Fig. 14 and accuracy of classification is given in Table VIII.

![Fig. 14. Classification result of 25 Bagged ionosphere dataset](image)

Table VIII
<table>
<thead>
<tr>
<th>Accuracy of Classification for Ionosphere Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy of bad return classification</td>
</tr>
<tr>
<td>Accuracy of good return classification</td>
</tr>
<tr>
<td>Overall dataset accuracy</td>
</tr>
</tbody>
</table>

The results in the 25 Bagged BPNN speak for themselves as the individual and overall efficiencies have increased greatly.

D. Classification of 25-Bagged ensemble dataset with PCA using BPNN

Now a modification is used in the 25 Bagged models that is dimensionality reduction. There are many methods for the purpose but the most vastly used in them is PCA (Principal Component Analysis). But instead of using PCA before bootstrapping we are using after that, this is a new approach but the results are promising. Also by doing this the diversity is increased between the bootstraps.

After applying the above model the output accuracies of both sonar and ionosphere datasets are given in Tables IX and X respectively.

![Fig. 15. The block diagram of 25 Bagged BPNN with PCA.](image)

Table IX
<table>
<thead>
<tr>
<th>Accuracy of Classification for Sonar Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy of mine classification</td>
</tr>
<tr>
<td>Accuracy of rock classification</td>
</tr>
<tr>
<td>Overall dataset accuracy</td>
</tr>
</tbody>
</table>

Table X
<table>
<thead>
<tr>
<th>Accuracy of Classification for Ionosphere Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy of bad return classification</td>
</tr>
<tr>
<td>Accuracy of good return classification</td>
</tr>
<tr>
<td>Overall dataset accuracy</td>
</tr>
</tbody>
</table>

Now in the case the accuracy decreased than the 25 Bagged simple BPNN but still it is greater than using a single NN.

VI. COMPARATIVE ANALYSIS

A. Sonar Dataset

The bar graph in Fig. 16 and Table XI show the comparison of sonar dataset accuracies taken from the above discussed models.

![Fig. 16. Comparison of sonar data accuracies](image)

Table XI
<table>
<thead>
<tr>
<th>Comparison of Sonar Dataset Accuracies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple NN</td>
</tr>
<tr>
<td>3-Bagged BPNN</td>
</tr>
<tr>
<td>25-Bagged BPNN</td>
</tr>
<tr>
<td>25-Bagged BPNN with PCA</td>
</tr>
</tbody>
</table>

The best accuracy for the sonar dataset is gained by using the ‘25-Bagged BPNN’. And so it is the most suitable, also the sonar dataset is not a huge dataset so the use of PCA is not necessary. The processing time for classification without PCA is in milliseconds and there is loss of almost 2% accuracy.
B. Ionosphere Dataset

The bar graph in Fig. 17 and Table XII show the comparison of sonar dataset accuracies taken from the above discussed models.

Table XII

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple BPNN</td>
<td>92.058</td>
</tr>
<tr>
<td>3-Bagged BPNN</td>
<td>89.4</td>
</tr>
<tr>
<td>25-Bagged BPNN</td>
<td>96.026</td>
</tr>
<tr>
<td>25-Bagged BPNN with PCA</td>
<td>94.039</td>
</tr>
</tbody>
</table>

Fig. 17. Comparison of ionosphere data accuracies

The best accuracy for the ionosphere dataset is gained by using the ‘25-Bagged BPNN’. The reasons are same as they are for the sonar dataset.

VII. CONCLUSIONS

- Bagging worked better than using a simple neural network for both sonar and ionosphere datasets.
- Also it was established that the number of bootstrap replicates must be significant enough for bagging to work particularly for small datasets.
- The number of bootstraps are not fixed they vary from dataset to dataset depending on the dimensionality, complexity and size of the data.
- Using PCA is useful only if accuracy is attained higher than the mere neural network.
- If application is not real time than the speed accuracy tradeoff is useless.

ACKNOWLEDGMENT

The authors would like to acknowledge their respective Universities and Institutes for providing adequate resources to conduct this research.

REFERENCES


Hafiz Tehzeebul Hassan did B.Sc. and M.Sc. in Electrical Engineering from UET, Lahore.

He later joined UET as a faculty member. His professional experience involves supervision of HV equipment tests as In-Charge of HV Lab in UET. He is an Associate Professor and a PhD student at UET Lahore. His research interests include Power System Analysis, Restructured Power Systems and Applications of Artificial Intelligence.

Mr. Hassan is a member of Pakistan Engineering Council.

Muhammad Umair Khalid did B.Sc. and M.Sc. in Electrical Engineering from UET, Lahore.

He is a HEC Scholar and a PhD student at UET Lahore. His research interests include Applications of Artificial Intelligence in Electrical Engineering.

Mr. Khalid is a member of Pakistan Engineering Council.

Kashif Imran received his elementary education from England where he obtained GCSEs and GCEs. He did B.Sc. and M.Sc. in Electrical Engineering from University of Engineering and Technology (UET), Lahore, Pakistan. His area of specialization during both degrees was electrical power.

He started his career as a Lecturer at UET Lahore. Then he moved to SIEMENS where he worked as Engineer on project coordination of 132kV grid stations. Later he joined NESPAX, leading engineering consultancy firm of Pakistan, as a Design Engineer in Power Distribution Section. In NESPAX, his professional experience includes design of overhead and underground power distribution systems for a variety of buildings and installations. Currently, he is Lecturer at COMSATS Institute of IT, Lahore, Pakistan where he teaches Electric Machines. He is an accomplished researcher with a high Impact Factor research paper titled “Simulation Analysis of Emissions Trading Impact on a Non-Utility Power Plant” that was published in Elsevier International Journal of Energy Policy in 2009. His book titled “Power Exchange as a Deregulated Electricity Market” is in press to be published by Lambert Academic Publishing, Germany. His research interests include Power System Economics, Restructured Power Systems Simulation, Micro-grid Energy Management Systems and Applications of Artificial Intelligence.

Mr. Imran is a member of Pakistan Engineering Council.