**Extensive Journal of Applied Sciences**
Available online at www.ejasj.com


# A study of Neural Networks for Classification: An Survey

**Ashish V Saywan**

***Corresponding author email:*** Ashish V Saywan

**ABSTRACT:** Classification is one of the most active research and application areas of neural networks. The literature is vast and growing. This paper summarizes the some of the most important developments in neural network classification research. Specifically, the issues of posterior probability estimation, the link between neural and conventional classifiers, generalization& learning tradeoff in classification, the feature variable selection, as well as the effect of misclassification costs are examined. Our purpose is to provide a synthesis of the published research in this area and stimulate further research interests and efforts in the identified topics. Index Terms—Bayesian classifier, classification, ensemble methods, feature variable selection, learning and generalization, misclassification costs, neural networks.

***Keywords*:** Neural Networks,Classification,survey.

## INTRODUCTION

classIfication is one of the most frequently encountered decision making tasks of human activity. A classification problem occurs when an object needs to be assigned into a predefined group or class based on a number of observed attributes related to that object. Many problems in business, science, industry, and medicine can be treated as classification problems. Examples include bankruptcy prediction, credit scoring, medical diagnosis, quality control, handwritten character recognition, and speech recognition.

Traditional statistical classification procedures such as discriminant analysis are built on the Bayesian decision theory .In these procedures, an underlying probability model must be assumed in order to calculate the posterior probability upon which the classification decision is made. One major limitation of the statistical models is that they work well only when the underlying assumptions are satisfied. The effectiveness of these methods depends to a large extent on the various assumptions or conditions under which the models are developed. Users must have a good knowledge of both data properties and model capabilities before the models can be successfully applied.

Neural networks have emerged as an important tool for classification. The recent vast research activities in neural classification have established that neural networks are a promising alternative to various conventional classification methods. The advantage of neural networks lies in the following theoretical aspects. First, neural networks are data driven self-adaptive methods in that they can adjust themselves to the data without any explicit specification of functional or distributional form for the underlying model. Second, they are universal functional approximators in that neural networks can approximate any function with arbitrary accuracy. Since any classification procedure seeks a functional relationship between the group membership and the attributes of the object, accurate identification of this underlying function is doubtlessly important. Third, neural networks are nonlinear models, which makes them flexible in modeling real world complex relationships. Finally, neural networks are able to estimate the posterior probabilities, which provides the basis for establishing classification rule and performing statistical analysis. On the other hand, the effectiveness of neural network classification has been tested empirically. Neural networks have been successfully applied to a variety of real world classification tasks in industry, business and science. Applications include bankruptcy prediction, handwriting recognition, speech recognition, product inspection, fault detection, medical diagnosis, and bond rating. A number of performance comparisons between neural and conventional classifiers have been made by many studies. In addition, several computer experimental evaluations of neural networks for classification problems have been conducted under a variety of conditions. Although significant progress has been made in classification related areas of neural networks, a number of issues in applying neural networks still remain and have not been solved successfully or completely.

In this paper, some theoretical as well as empirical issues of neural networks are reviewed and discussed. The vast research topics and extensive literature makes it impossible for one review to cover all of the work in the filed. This review aims to provide a summary of the most important advances in neural network classification. The current research status and issues as well as the future research opportunities are also discussed. Although many types of neural networks can be used for classification purposes, our focus nonetheless is on the feed forward multilayer networks or multilayer perceptron's (MLPs) which are the most widely studied and used neural network classifiers. Most of the issues discussed in the paper can also apply to other neural network models.

The overall organization of the paper is as follows. After the introduction, we present fundamental issues of neural classification in Section II, including the Bayesian classification theory, the role of posterior probability in classification, posterior probability estimation via neural networks, and the relationships between neural networks and the conventional classifiers. Section III examines theoretical issues of learning and generalization in classification as well as various practical approaches to improving neural classifier performance in learning and generalization. Feature variable selection and the effect of misclassification costs—two important problems unique to classification problems—are discussed in next few Sections, respectively.

## *NEURAL NETWORKS AND TRADITIONAL CLASSIFIERS*

### *A. Bayesian Classification Theory:*

Bayesian decision theory is the basis of statistical classification methods. It provides the fundamental probability model for well-known classification procedures such as the statistical discriminant analysis.

Consider a general -group classification problem in which each object has an associated attribute vector of dimensions. Let denote the membership variable that takes a value of if an object is belong to group. Define as the prior probability of group and as the probability density function.

### *B. Posterior Probability Estimation via Neural Networks:*

In classification problems, neural networks provide direct estimation of the posterior probabilities. The importance of this capability is summarized by Richard and Lippmann:

"Interpretation of network outputs as Bayesian probabilities allows outputs from multiple networks to be combined for higher level decision making, simplifies creation of rejection thresholds, makes it possible to compensate for difference between pattern class probabilities in training and test data, allows output to be used to minimize alternative risk functions, and suggests alternative measures of network performance."

A neural network for a classification problem can be viewed as a mapping function, where input is submitted to the network and an -vectored network output is obtained to make the classification decision. The network is typically built such that an overall error measure such as the mean squared errors (MSE) is minimized. From the famous least squares estimation theory in statistics, the mapping function which minimizes the expected squared error.

### *C. Neural Networks and Conventional Classifiers:*

Statistical pattern classifiers are based on the Bayes decision theory in which posterior probabilities play a central role. The fact that neural networks can in fact provide estimates of posterior probability implicitly establishes the link between neural networks and statistical classifiers. The direct comparison between them may not be possible since neural networks are nonlinear model-free method while statistical methods are basically linear and model based.

As the statistical counterpart of neural networks, discriminant analysis is a well-known supervised classifier. Gallinari at all. The discriminating rule is simply: assign describe a general framework to establish the link between discriminant analysis and neural network models. They find that in quite general conditions the hidden layers of an MLP project the input data onto different clusters in a way that these clusters can be further aggregated into different classes. For linear MLPs, the projection performed by the hidden layer is shown theoretically equivalent to the linear discriminant analysis. The nonlinear MLPs, on the other hand, have been demonstrated through experiments the capability in performing more powerful nonlinear discriminant analysis. Their work helps understand the underlying function and behavior of the hidden layer for classification problems and also explains why the neural networks in principle can provide superior performance over linear discriminant analysis. The discriminant feature extraction by the network with nonlinear hidden nodes has also been demonstrated in Asoh and Otsu and Webb and Lowe. Lim, Alder and Had Ingham show that neural networks can perform quadratic discriminant analysis.

Raudys presents a detailed analysis of nonlinear single layer perceptron (SLP). He shows that during the adaptive training process of SLP, by purposefully controlling the SLP classifier complexity through adjusting the target values, learning-steps, number of iterations and using regularization terms, the decision boundaries of SLP classifiers are equivalent or close to those of seven statistical classifiers. These statistical classifiers include the Euclidean distance classifier, the Fisher linear discriminant function, the Fisher linear discriminant function with pseudo-inversion of the covariance matrix, the generalized Fisher linear discriminant function, the regularized linear discriminant analysis, the minimum empirical error classifier, and the maximum margin classifier. Kanaya and Miyake and Miyake and Kanaya also illustrate theoretically and empirically the link between neural networks and the optimal Bayes rule in statistical decision problems.

Logistic regression is another important classification tool. In fact, it is a standard statistical approach used in medical diagnosis and epidemiologic studies. Logistic regression is often preferred over discriminant analysis in practice .In addition, the model can be interpreted as posterior probability or odds ratio. It is a simple fact that when the logistic transfer function is used for the output nodes, simple neural networks without hidden layers are identical to logistic regression models. Another connection is that the maximum likelihood function of logistic regression is essentially the cross-entropy cost function which is often used in training neural network classifiers. Schumacher *et al.* make a detailed comparison between neural networks and logistic regression. They find that the added modeling flexibility of neural networks due to hidden layers does not automatically guarantee their superiority over logistic regression because of the possible over fitting and other inherent problems with neural networks.

### *LEARNING AND GENERALIZATION*

Learning and generalization is perhaps the most important topic in neural network research. Learning is the ability to approximate the underlying behavior adaptively from the training data while generalization is the ability to predict well beyond the training data. Powerful data fitting or function approximation capability of neural networks also makes them susceptible to the over fitting problem. The symptom of an over fitting model is that it fits the training sample very well but has poor generalization capability when used for prediction purposes. Generalization is a more desirable and critical feature because the most common use of a classifier is to make good prediction on new or unknown objects. A number of practical network design issues related to learning and generalization include network size, sample size, model selection, and feature selection. Wolpert addresses most of these issues of learning and generalization within a general Bayesian framework.

In general, a simple or inflexible model such as a linear classifier may not have the power to learn enough the underlying relationship and hence under fit the data. On the other hand, complex flexible models such as neural networks tend to over fit the data and cause the model unstable when extrapolating. It is clear that both under fitting and over fitting will affect generalization capability of a model. Therefore a model should be built in such a way that only the underlying systematic pattern of the population is learned and represented by the model.

The under fitting and over fitting phenomena in many data modeling procedures can be well analyzed through the well-known bias-plus-variance decomposition of the prediction error. In this section, the basic concepts of bias and variance as well as their connection to neural network classifiers are discussed. Then the methods to improve learning and generalization ability through bias and/or variance reductions are reviewed.

A. Bias and Variance Composition of the Prediction

B. Methods for Reducing Prediction Error

## CONCLUSION

Classification is the most researched topic of neural networks. This paper has presented a focused review of several important issues and recent developments of neural networks for classification problems. These include the posterior probability estimation, the link between neural and conventional classifiers, the relationship between learning and generalization in neural network classification, and issues to improve neural classifier performance. Although there are many other research topics that have been investigated in the literature, we believe that this selected review has covered the most important aspects of neural networks in solving classification problems.

The research efforts during the last decade have made significant progresses in both theoretical development and practical applications. Neural networks have been demonstrated to be a competitive alternative to traditional classifiers for many practical classification problems. Numerous insights have also been gained into the neural networks in performing classification as well as other tasks. However, while neural networks have shown much promise, many issues still remain unsolved or incompletely solved. As indicated earlier, more research should be devoted to developing more effective and efficient methods in neural model identification, feature variable selection, classifier combination, and uneven misclassification treatment. In addition, as a practical decision making tool, neural networks need to be systematically evaluated and compared with other new and traditional classifiers. Recently, several authors have pointed out the lack of the rigorous comparisons between neural network and other classifiers in the current literature .This may be one of the major reasons that mixed results are often reported in empirical studies.

Other research topics related to neural classification include network training, model design and selection, sample size issues, Bayesian analysis, and wavelet networks. These issues are common to all applications of neural networks and some of them have been previously reviewed. It is clear that research opportunities are abundant in many aspects of neural classifiers. We believe that the multidisciplinary nature of the neural network classification research will generate more research activities and bring about more fruitful outcomes in the future.

## REFERENCES

K. Al-Ghoneim and B. V. K. V. Kumar, "Learning ranks with neural networks," in Proc. SPIE Applications Science Artificial Neural Networks, 1995.

E. I. Altman, G. Marco, and F. Varetto, "Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience)," J. Bank. Finance, 1994.

B. Amirikian and H. Nishimura, "What size network is good for generalization of a specific task of interest?," Neural Networks, 1994.

U. Anders and O. Korn, "Model selection in neural networks," Neural Networks, 1999.

N. P. Archer and S. Wang, "Fuzzy set representation of neural network classification boundaries," IEEE Trans. Syst., Man, Cybern., 1991.

L. Atlas, R. Cole, J. Connor, M. El-Sharkawi, R. J. Marks II, Y. Muthusamy, and E. Barnard, "Performance comparisons between backpropagation networks and classification trees on three real-world applications," in Advances in Neural Information Processing Systems, D. S. Touretzky, Ed. San Mateo, CA: Morgan Kaufmann, 1990.

A. Babloyantz and V. V. Ivanov, "Neural networks in cardiac arrhythmias," in Industrial Applications of Neural Networks, F. F. Soulie and P. Gallinari, Eds, Singapore: World Scientific, 1998.

P. F. Baldi and K. Hornik, "Learning in linear neural networks: A survey," IEEE Trans. Neural Networks, vol. 6, 1995.

E. Barnard, "Optimization for training neural nets," IEEE Trans. Neural Networks, 1992.

E. Barnard and E. C. Botha, "Back-propagation uses prior information efficiently," IEEE Trans. Neural Networks, 1993.

R. Barron, "Statistical properties of artificial neural networks," in Proc. 28th IEEE Conf. Decision Control, 1989.

R. Battiti, "First-and second-order methods for learning; between steepest descent and Newton's method," Neural Comput., 1992.

"Using mutual information for selecting features in supervised neural net learning," IEEE Trans. Neural Networks, 1994.

R. Battiti and A. M. Colla, "Democracy in neural nets: Voting schemes for classification," Neural Networks, 1994.

E.B.Baum, "What size net gives valid generalization?," Neural Computer., 1989.

J. M. Benitez, J. L. Castro, and I. Requena, "Are artificial neural networks black boxes?," IEEE Trans.