CHAPTER 2

REVIEW & LITERATURE

2.1 Introduction

In the following paragraphs, some of the studies related to the subjects of our research work have been discussed and which have been previously undertaken in the related fields of predictive data mining by traditional and new techniques, decision support systems, predictive data mining, building predictive models with artificial neural networks, first and second order gradient based algorithms for training the neural networks and the insurance sector. In recent years, a lot of work has already been done in this field that includes employing intelligent methods for developing the predictive models and decision making. Empirical comparisons have also been investigated by researchers using different data mining techniques. A brief review on the research work done in the above mentioned fields has been given here and the reviews have been classified according to the subject in the review headings.

2.2 Reviews on Use of Artificial Neural Networks in Decision Support Systems

In chapter 26 ‘Artificial Neural Networks in Decision Support Systems’ of the book ‘Handbook on Decision Support Systems 1: Basic Themes’ written by Delen and Sharda, authors have introduced the basic concepts of neural networks and have presented an overview of the possible applications of neural networks based decision support systems (DSS). According to authors, ANNs can be viewed as supporting at least two different types of decision support systems: data driven and model-driven systems. On one side, ANNs can be used as data analysis tools for forecasting and prediction based on historical data in a data-driven DSS. On the other side, ANNs can also be viewed as a class of quantitative models that can be used in a model-driven DSS. After describing the basic concepts of neural networks in the chapter, they have presented some selected applications of neural networks in DSS. Authors have introduced about their research on a web-based decision support system that employs a neural network. This system has been built to help a Hollywood decision maker in
making decisions on certain movie’s parameters. The chapter also discusses about a list of issues that should be considered in employing a neural network for any DSS application.

The authors mention that the increasing uncertainty and complexity associated with today’s DSS created a need to apply sophisticated quantitative models that have the capabilities beyond traditional simple linear models. Since the uncertainty and complexity among the data (that describes decision situations) are increasing, therefore the capabilities of the models (that represents these kinds of situations) should also increase, so that highly non-linear relationships among the variables can also be functionalized. This is why the artificial neural networks (ANNs) fit into the realm of managerial decision support systems (DSS).

In this chapter the author has explained about some important applications of neural networks in decision support systems (DSS). He has discussed about Medical DSS applications of neural networks, Financial DSS applications of neural networks and some other important DSS applications of neural networks including environmental issues. He has also discussed about self-developed DSS as an exemplary case: A Web-Based DSS for Hollywood Managers.

Power⁴³ has discussed about the classification decision support systems in his research and according to him ANNs can be considered to play an important role of the “quantitative models” in the model-driven DSS. Power introduced five different categories based on various dominant components of the DSS, data/document-driven, knowledge-driven, communication or/ group-driven model-driven and web-based or/inter-organizational DSS. According to him Model-driven DSS refers to the type of decision support systems, where an underlying model of a specific situation is constructed and it is then used for analysis of different alternatives and aid in decision making. Traditionally, these models have been categorized in many different types: optimization, simulation, decision analysis, etc.⁴⁴ According to the author neural networks would now represent another type of modeling approach. Neural networks could now be considered to relate to the data-driven decision support systems. Neural networks (ANNs) provide a method for forecasting and analyzing the historical data trends. Integration of model-driven DSS with other DSS technologies like expert systems, artificial intelligence models, artificial neural networks etc. are also an active
area of research. Authors have researched their own DSS and their model-driven DSS called Movie Forecast Guru allows a user to generate the forecast for a movie’s box office performances by changing the movie’s decision input parameters like (star power, season of release, special effects, etc.). Changes in these predictor parameters have been analyzed by a number of trained forecasting models based upon methods such as artificial neural networks, logistic regression and discriminate analysis.

Tsadiras, Papadopoulos and Kelly⁴⁵ in their research paper “An artificial neural network based decision support system for solving the buffer allocation problem in reliable production lines” have presented the study about development of ANN based DSS. In their research work, they have tried to solve one of the major design problems in the area of manufacturing systems and the problem is widely known as Buffer Allocation Problem (BAP). More precisely, an ‘Artificial Neural Network based Decision Support System’ (ANN DSS) has been developed by them to support production line engineers in making the decisions concerning to Buffer Allocation Problem (BAP) in reliable production lines. The aim of the ANN DSS is to predict about the performance of the production line based on its input characteristics. The decision support system has been planned to allow for these predicted data to be outputted in a user friendly and intelligent manner.

The performance of the system has been examined for test sets of production lines and an average accuracy near to 99% has been achieved. The performance of the ANN based DSS model has been compared with that of other traditional and well established surface fitting techniques and its superiority has been confirmed.

The developed ANN based decision support system, called BAPANN DSS, has been designed and implemented in MATLAB Neural Network Toolbox. BAPANN’s functionalities and capabilities have been tested in different scenarios, to prove the effectiveness of the proposed methods and are measured in terms of the CPU time required to train the network and solution of problem. According to authors, BAPANN DSS has proved it a powerful, efficient and accurate tool to make decisions on the ‘buffer allocation problem’ (BAP) and is used for balanced reliable production lines. Also the complex problem has been solved in a convenient manner without involving the engineers in tedious and complex mathematical analysis of the problem.
Conrads and Roehl Jr.\textsuperscript{46} have mentioned about development of Decision Support Systems based on Artificial Neural Networks in their research paper “An Artificial Neural Network-Based Decision Support System to Evaluate Hydropower Releases on Salinity Intrusion” presented in 7th International Conference on Hydroinformatics, France.

The ANN models developed by them are capable to convincingly reproduce historical behaviors and are able to generate alternative scenarios of interest. To make the ANN models directly available to all stakeholders, an intelligent user-friendly decision support system DSS has been developed and it integrates together the historical databases, ANN based DSS models, all user controls, streaming graphics and simulation outputs.

For the research, individual ANN models have been developed for nine continuous coast stream gages. The models have been developed in two different stages. The first stage models the chaotic, lower-frequency, portion of signal, as represented by the filtered SC signals. The second stage models the periodic, high-frequency, hourly SC, using the predicted SC as a carrier signal. The model in the DSS has been called a “super model” that represents the whole system. The super model is made up of the 18 “sub-models” of the daily and hourly models for each gaging station.

Arsene, Gabrys and Al-Dabass\textsuperscript{47} in their research paper “Decision support system for water distribution systems based on neural networks and graphs theory for leakage detection” have discussed about the development and testing of an effective and efficient decision support system (DSS) for operational monitoring and control of water distribution systems based on a three-layer General Fuzzy Min–Max Neural Network GFMMNN and graph theory.

The operational monitoring and control of the system involves detection of pipe leakages. The training data for the GFMMNN DSS has been obtained through simulation of leakages in a water network for a 24 hour operational period. A combination of three numerical algorithms for modeling and simulation of water networks have been based on loop ANN and graph theory. It has been shown that the detection of leakages based on the training and testing of the GFMMNN model with patterns of variation of nodal consumptions with or without the confidence limits
outputs better recognition rates in comparison to the training based on older technique of patterns of nodal heads and pipe flows state estimates with or without confidence limits. The model depicts comparable recognition rates to the original recognition system trained with patterns of data obtained with the LS nodal heads state estimator.

Bhattacharya, Xuand Kumar\textsuperscript{48} have carried out the research work on the development of ANN based DSS for financial fraud detection and have explained about it in their research paper “An ANN-based auditor decision support system using Benford’s law”. In their research they have observed that there is a growing professional interest on the application related to financial fraud detection, but there has been relatively little academic research to demonstrate its efficacy as a decision support tool and in the context of an analytical review procedure pertaining to a financial audit. They have conducted a numerical study using an optimized artificial neural network ANN. Building on an earlier work by others of a similar nature; they have assessed the benefits of ANN as a useful classifier in segregating naturally occurring numbers. They have come up with the optimal network structure for every instance corresponding to a 3×3 Manipulation–Involvement matrix which has drawn to depict the different combinations of the level of sophistication in data manipulation by the agents of a financial fraud and also the extent of collusive involvement.

Bhande et al. have suggested the development of an Intelligent Decision Support System (IDSS) for early prediction of Parkinson disease with help of soft computing techniques.\textsuperscript{49} In their research work techniques have been proposed based on artificial neural networks (ANNs) and older statistical methods have also been adopted to help the specialists in the more accurate diagnosis of Parkinson disease. Data have been recorded in 195 examinations and carried out on 31 patients; and the dataset are used to confirm the capacity of the suggested binary classifier system. The paper discusses the techniques used to diagnose the disease in general. From exhaustive and careful experimentation, it is concluded that multilayer perceptron neural network based classifier ensures true estimation of the complex decision boundaries, remarkable discriminating capacity and displays better results than other architectures. The research work shows that the percentage average classification accuracy of multilayer perceptron model based on ANN is better as compared to other techniques and it has achieved very good results on such complex and insufficient data set. The
classification ability of the multilayer perceptron neural network (MLPNN) model is not only better than rule based statistical models, but also better than other NN based models as RBF and PCA. It has also been found that the MLPNN based DSS deliver consistently good performance. Hence, the suggested MLPNN based DSS can be employed as a good decision-making aid by the physicians and it can always be applied for meticulous cross-examination and can serve as a second opinion for enhanced diagnosis of the disease.

2.3 Reviews on Predictive Data Mining

Many researchers have applied various techniques like traditional statistical techniques, regression, decision tree, naive bayes and the new techniques like soft computing based methods - neural networks, genetic algorithms, fuzzy techniques, evolutionary algorithms and hybrid techniques to analyze quantitative data for the forecasting purposes. The research has been applied particularly in the fields of predictive data mining for business applications, medical field like disease diagnosis and various important engineering fields etc.

Han and Kamber in their book on “Data Mining: Concepts and Techniques”\textsuperscript{24}, U. M. Fayyad et. al. in their book on “Advances in Knowledge Discovery and Data Mining”\textsuperscript{50}, Kargupata and Joshi: in their book on “Data Mining: Next Generation Challenges and Future Directions”\textsuperscript{51} discuss the importance of creating data warehouses and predicting future trends based upon historical trends present among the data for decision making and have explained the basic concepts of data mining and their use in real life applications. They talk about all possible kinds of data mining functionalities and methods for the implementation. Authors have explained about the classification and prediction techniques, difference among the two techniques and how these techniques can be applied to real life situations. Various methods and algorithms for different categories of data mining including classification and prediction have been explained in detail. Techniques like linear regression, non-linear regression and decision tree based methods, error back propagation with ANN, statistical methods of estimation of accuracy and error measures for the analysis have been discussed. This book provides a lot of algorithms which are geared to the discovery of predictive data patterns hidden in large, real datasets and the pseudo-codes have been presented for the algorithms which can easily be translated into popular
programming languages of research. The book explores in detail the concepts and techniques of predictive data mining, which is a promising and nourishing frontier in database systems and upcoming new data mining applications.

According to Guo\textsuperscript{52}, predictive data mining combines procedures from fields of machine learning, pattern recognition, statistics, database theory and visualization to extract the concepts, inter-relations and interesting patterns automatically from a huge corporate databases. The primary goal is to extract knowledge from data to support the decision-making process. Two primary functions of data mining are: prediction, which involves finding unknown values/relationships/patterns from known values; and description, which provides interpretation of a large database. In his research paper author introduces predictive data mining based upon Decision Trees and Logistic Regression. He uses CART and CHAID techniques of decision trees to classify a data set. He has examined numerous risk-factors for automobile drivers insurance with the goal of predicting their claim frequency. The influence and the correlation of these factors on auto claim distribution has been identified with exploratory data analysis and decision tree algorithm. Logistic regression has been applied to model claim frequency.

According to Mittal and Dhingra\textsuperscript{53}, mining of predictive information which is hidden in large databases of companies is a powerful upcoming technology and this has a great potential to help companies and can focus on the most important information in their data warehouses. Various upcoming data mining tools have the ability to predict future trends and patterns and allow the businesses to make proactive and knowledge-based decisions. Authors have also discussed about different data mining functionalities and introduced the new data mining techniques like artificial neural networks, decision trees, genetic algorithms, link analysis, nearest neighborhood methods for the data mining purpose. In the research papers, they have also focused upon various applications possible with data mining and highlighted the future trends possible in the fast emerging field.

Berson and Smith\textsuperscript{7} in their book titled “Data Warehousing, Data Mining and OLAP” have mentioned about the use of various data mining techniques and technologies for the decision support and talk about future trends and analytics. They talk about traditional methods in use for the purpose and newly upcoming methods
like artificial intelligent techniques based technologies like development of expert systems, OLAP, fuzzy logic based algorithms, genetic algorithms, neural networks based techniques etc. In chapter 19, the authors have given a brief review of how artificial neural networks can be employed for the predictive data mining. They also discuss about different kinds of real world situations, where ANN can be applicable and discuss about their relative score cards. In the same chapter case study depicting the use of ANN in business application of predicting the currency exchange rates is provided and also focused on the strengths and weaknesses of employing the ANNs for prediction purpose. According to the authors ANNs are relatively fast to apply to the data although they do require significant numerical computations. They are highly scalable in nature and can be applied to very large databases as well as can be trained in parallel on SMP and MPP supercomputers.

Mishra, Das, Mausumiet. al.\textsuperscript{54} in their research paper “Predictive Data Mining: Promising Future and Applications” have focused on the important traditional predictive analytics techniques and important application areas for these techniques. According to authors Predictive analytics is a branch of data-mining which is concerned with the prediction of future probabilities and trends. The central elements of predictive analytics are the predictor or input variables which can be measured for an individual or other entity to predict or output future behavior. An insurance company for example, may likely to take into account potential safety predictors of driving such as gender, age, driving record etc. when insuring policies for car insurance. Multiple predictors can be further combined into a predictive model (can be based on a variety of techniques), which, when subjected to the analysis, can be used to predict the future probabilities with an adequate level of reliability. In predictive data mining models, data are collected, a statistical model is formulated, predictions are made using the model and the developed model is validated (or revised) as new data becomes available. Authors also mention the important application areas and explain that predictive analytics can be applied to many research areas, security, genetics, economics, marketing and other engineering fields. In this paper, authors have discussed aboutand done extensive research on various predictive techniques with all its future directions and applications in various areas.
Traditional statistical techniques of predictive data mining like linear regression, multiple linear regression, logistic regression and application areas of these techniques have been explained in the research paper. This research mainly intends to concentrate on the mining techniques using predictive data mining. Since predictive analytics is a major area of interest to almost all communities and organization, with a vast range of application areas. And at the same time the widespread availability of several new upcoming computational tools and technologies for predictive modeling assists the researchers and the practitioners to select the most appropriate strategy for their models. Authors have mentioned some important software tools that enable greater transparency and are capable to analyze past trends to predict the probable future outcome of an event or its likelihood to occur, as well as to discover the hidden patterns and relations among historical data.

2.4 Reviews on Predictive Data Mining with Artificial Neural Networks

Ni in his research paper explains about the process of developing prediction models based on artificial neural networks. He discusses about application with feed-forward and ART based neural networks. He explains that although neural networks have complex internal structure, can take long training time and its internal computations are not easily understandable, but neural networks possess high acceptance ability for noisy and non-linear data and show high accuracy in results and therefore are preferable in data mining. The author says that at present, data mining is a new and developing area of research and neural networks are very suitable for solving the problems of data mining due to their characteristics of good robustness, self-organizing, adaptive, parallel processing, support distributed storage and exhibit high degree of fault tolerance. The combination of data mining methods and neural network architecture based models can greatly enhance the efficiency of data mining techniques.

Patel and Goyal suggest that Artificial Neural Networks (ANNs) can be applied to a variety of real-life problems where relationships between input and output variables are complex in nature. ANNs can be employed to find solution for the problems that are not simple to solve with conventional technologies, where finding of algorithmic
solutions seems to be difficult. Since their modeling is inspired from biological neural networks present in human brain, they are well suited for problem solving that is suitable for human brain, but for which computers in general are not. They are found to be suitable for problems like pattern recognition and prediction. The later techniques require the recognition of future trends based upon trends and relationships present among historical data. ANNs are also found suitable for problems involving adaptive learning (means capability to learn and perform tasks based on earlier experiences) and self-organization (means artificial neural networks have the capability to create their own organization or representation of the information which has been received during training time and real-time operations).

Tezel and Buyukyildizin the research paper “Monthly evaporation forecasting using artificial neural networks and support vector machines” have focused on usability of artificial neural networks (ANNs) multilayer perceptron (MLP) networks, radial basis function networks (RBFN) and support vector machines (SVM) to forecast about monthly evaporation of the hydrological cycle. According to authors evaporation has been one of the most important components of the hydrological cycle and is relatively difficult to estimate, due to its complexity, because it can be influenced by a large number of factors. Estimation of evaporation is very crucial for the quality design of reservoirs, especially in arid and semi-arid areas. Artificial neural network methods and support vector machines (SVM) based predictive models are utilized to estimate evaporation and other hydrological variables. In this study, usability of artificial neural networks (ANNs) (multilayer perceptron (MLP) and radial basis function network (RBFN)) and support vector regression (SVR) artificial intelligence methods have been investigated to estimate the monthly pan-evaporation. For this measurement - temperature, relative humidity, wind speed and precipitation data for the period 1972 to 2005 from Beysehir meteorology station have been used as input (predictor) variables while pan evaporation values were used as output. Earlier popular techniques like Romanenko and Meyer method have been also considered for the comparison. For ANN based MLP architecture, four different training algorithms, gradient descent with momentum and adaptive learning rule back propagation (GDX), Levenberg–Marquardt (LVM), scaled conjugate gradient (SCG) and resilient back propagation (RBP), have been applied. The models have been designed via 10-fold cross-validation (CV); algorithm performance has been assessed via mean absolute
error (MAE), root mean square error (RMSE) and coefficient of determination ($R^2$). According to the experimental observations, the ANN algorithms and SVR had similar performance results. The ANNs and SVR methods have been found to perform better than traditional methods. Consequently, the best performance using the test data has been obtained using SCG(4,2,2,1) neural network structure, with $R^2=0.905$. The results obtained by using the data belonging to Beysehir observation station have shown that artificial intelligence (AI) based new methods such as ANN and SVR have achieved significant success in the approximation of air evaporation amount function.

Jayaraj and Raman have proposed a work on the topic “An hybrid multilayer perceptron using GSO-GA for software defect prediction” in the year 2016. As we know that software defect prediction has turned into an expected requirement for organizations to guarantee quality and reliability of software products. In this work they oversee software defects as 3 levels (defect discovery, defect analysis, defect prevention). Software defect prediction (SDP), procedures are utilized either to classify which modules are defect inclined or to predict the quantity of defects anticipated that would be found in a software module. So to overcome these defects they has proposed a hybrid swarm optimization, genetic algorithm, ANN based model and found that this algorithm has better predictive values and the datasets utilized has been taken from NASA(MDP)-metric data repository. From experimental observations it can be clearly concluded that the new hybrid algorithm beats the conventional algorithms in performance and the outcomes demonstrated that the algorithm has shown enhanced classification accuracy and precision, values, when compare to the other methods.

Kutylowska in his research paper “Neural network approach for failure rate prediction” has presented the possibility of artificial neural networks (ANN) application to the failure rate modeling. Operating data from one Polish water utility has been used to predict output values for the failure frequency. The prediction results clearly indicate that artificial neural networks (ANNs) can be successfully used to model the damaged frequency in the water supply systems. It has been found that the artificial neural network multilayer perceptron (MLP) architecture trained by quasi-Newton approach has given acceptable performance and convergence from engineering
point of view. The network has been trained using 173 and 147 data (house connections and distribution pipes, respectively). 50% of all dataset has been designated for training, 25% for testing and 25% for validation. In the prognosis phase, the best created network has used 100% of 133 and 114 values for testing. The correlation between experimental and predicted data (relating to house connections and distribution pipes, respectively) has been characterized by indicator $R^2 = 0.9510$ and $R^2 = 0.9268$ (learning phase). In this step of modeling once created network predicted failure rate using not known input signals. The coefficient $R^2$ was equal to 0.4142 for house connections. The created model could be successfully used by water utility in the future to find the level of failure frequency and to thus to plan the necessary renovation of the most deteriorated pipes.

2.5 Reviews on Related Works Done with ANN and other Techniques in Medical Field for Disease Diagnoses

Søreide, Thorsen and Søreide in their research work have developed prediction models for patients with perforated peptic ulcer (PPU). In their research they observed that traditional techniques have failed to produce consistent results with higher accuracy because of complex nature of the disease and the data exhibit many non-linear associations with outcomes. But during their experimentations they have explored that artificial neural networks (ANNs) have the capabilities to predict and functionalize the complex interactions present between the risk factors of PPU and death among patients suffering with this disease. They utilized for their research work on a set of 172 patients, 168 had their data involved in the model; the data of 117 patients (70 %) were utilized in the training of model and the data of 51 patients (39 %) were utilized for the testing the models developed in IBM SPSS. This research demonstrated that use of ANNs to enhance the outcome predictions regarding patients with PPU has been successful and accuracy has been slightly improved in comparison to the predictive abilities of a traditionally applied statistical regression model.

Gharehchopogh et al. have done the case study in diagnosis of thyroid disease with help of artificial neural networks. The idea behind the use of ANNs is to diagnose the disease and to increase the accuracy of performance. The appropriate selection of neural architectures affects the network performance in terms of convergence and to
reach the point of higher accuracy. By selecting a hidden layer and log-sigmoid activation function for hidden layer and 6 neurons in the hidden layer, they were able to reach the classification accuracy for thyroid disease to 98.6%. The proposed method in this paper can be a solution to increase the performance of ANN based prediction models. So, it can also be generalized for other disease diagnoses systems or decision support systems or engineering applications.

Sibanda and Pretorius have applied Multilayer Perceptron Neural Networks (MLPNN) to Model HIV in South Africa. In their research work they have presented the use of multilayer perceptron neural networks (MLPNN) to model the demographic characteristics of antenatal clinic attendees in South Africa. MLPNN have been employed to develop predictive models to classify both the HIV negative and positive clinic attendees and simulations developed have been assessed using validity and reliability of the tests. According to them neural networks are found to be robust in sampling variations and for classification performance. MLPNNs prove to be a useful tool for data mining functionalities like prediction, classification and function approximation. They further emphasize that practical benefits of employing neural networks in modeling predictive systems are huge. In their research they present the performance and MSE functions with respect to different parameters like numbers of hidden layers and number of iterations etc. They have observed that performance increases with the number of neurons in the hidden layers and it is found to be 66% prediction with one neuron, 69%, 71%, 72% and 74% prediction respectively for two, four, five and ten hidden layers.

Use of Naive Bayes, Artificial Neural Networks and Decision Trees methods has been proposed by Palaniappan et al. An Intelligent Heart Disease Prediction System (IHDPS) has been developed by using the above said data mining techniques. Each method has its own pros and cons to get appropriate results. To build the system and to extract the hidden knowledge and relationship between predictor variables have been used.

Decision Tree based techniques include CART (Classification and Regression Tree), ID3 (Iterative Dichotomized 3) and C4.5. These methods differ in selection of splits, when to stop a node from splitting and assignment of class to a non-split node. CART makes use of Gini index to measure the impurity of a partition or set of training
tuples. Decision Trees are able to handle continuous data (as in regression) but they must be converted to categorical data. According to authors Naive Bayes method is the basis for many machine-learning and traditional data mining techniques. The rule (algorithm) can be used to create models with predictive capabilities and can provide new ways of understanding and exploring data. It learns from the “evidence” by calculating the correlation between the variables. In ANN based models connection between input units and hidden and output units are based on relevance of the assigned value (weight) of that particular input unit. A total of 909 records with 15 medical attributes (factors) have been applied from publically available Cleveland Heart Disease database.

According to Bhatla and Jyoti different data mining methodologies can be employed in automated heart disease prediction systems and in recent time various data mining techniques have been emerging for effective and efficient heart disease prediction and diagnosis. The analysis shows that neural network with 15 attributes has shown the higher accuracy. On the other hand, decision tree has also performed well with but less accuracy than neural network, by using 15 attributes. Moreover, in combination with genetic algorithm and 6 attributes, decision tree has shown and hybrid techniques also show sufficiently good efficiency.

In the research work proposed by Nayeem Wan and Hasan, ANN and their ability to learn complex and non-linear relationships including noisy or less precise information, neural networks have been well suited to solve problems in the medical field. The feed-forward MLP with back propagation neural network and with supervised learning techniques have been proposed to diagnose and predict the disease. MLPNN architecture has been used to train the neural network model for the prediction purpose. Their proposed network model shows an accuracy of 84% for heart disease prediction where the existing and traditional models exhibit accuracy of only 82%. This is happened because of the use of increased hidden layers, more neurons in hidden layer and use of more training data. The authors of the research paper have also tested the model for liver disorder diagnosis and lung cancer prediction and the same model networks shows an accuracy of 82% and 91% respectively.
Awang and Siraj in their research work intent to apply artificial neural network in predicting the occurrence of heart disease, mainly the angina in patients. The prediction and diagnosis of angina disease are important to determine the most suitable form of treatment for these types of patients. The development of this application has been done in three main phases. In the first phase development of Heart Disease Management Information System (HDMIS) for data collection and patient management has been done. The second phase focuses upon the development of a neural network model based on back propagation technique neural network for training and testing. The final phase concentrates on the evolution of a prediction model and to apply the research for prediction on new patients’ data. The best observation achieved for the prediction accuracy of 88.89 %. Apart from proving the capability of artificial neural networks in field of medical diagnosis, the study also describes that neural network based prediction can be incorporated into the hospital information and decision support systems. As the pilot project, the research explains how ANNs can be utilized as a medical decision support system and can be applied particularly in diagnosis of the heart disease.

According to Rani classification and prediction is an important problem and their use is one of the fast emerging field of data mining. Neural networks are very much suitable to solve problems in the field of biomedical engineering. In her study and research, neural network modeling technique is adopted for prediction modeling of medical dataset. The experiment is performed with Heart Disease dataset and implementation of the single and multilayer neural network modes has been done. Back propagation algorithm with learning rate momentum and variable learning rate has been used and tested to train the neural networks.

Chowdhury et al. in their research work show the use of artificial neural networks (ANNs) for prediction of neonatal disease and its diagnosis. The proposed methods involve training a MLPNN model with back propagation learning to predict various categories of neonatal diseases. 94 simulations have been tested using this model. The research finds ANN based prediction of neonatal disease has prediction accuracy of 75%.

Kaur and Singh in their survey and review on heart disease prediction system using data mining techniques has explained about the possible data mining techniques and
variety of tools available for the prediction of heart disease. They have discussed about the traditional and newly emerging technologies in the field of prediction like use of artificial neural networks. A comparative table has also been discussed showing results of different data mining techniques used in the diagnosis of heart disease and applied over different heart disease datasets. Various data mining techniques and classifiers are discussed and defined in the research; which have been emerging in the past years for an efficient and effective heart disease prediction and diagnosis. The analysis shows that different tools and technologies have been used in all the papers by various authors while considering different number of attributes.

Kumari et al. focuses on solution of heart disease datasets utilizing different types of data mining algorithms such as: Artificial neural networks (ANNs), Decision Trees, Support Vector Machine (SVM). The research analyses the performance factor employing these algorithms through several statistical methods and newly evolving methods. Accuracy of Decision Trees, ANNs and SVMs are found to be 79.05%, 80.06% and 84.12% respectively. Error rates for Decision Trees, ANNs and SVMs are found to be 2.756, 0.2755, 0.2248 and 0.1588 respectively. The analysis shows that out of these four classification models SVMs and ANNs predict heart disease with least error rate and highest accuracy.

According to Wadhonkar, Tijare and Sawalkar, Artificial Neural Networks (ANNs) are gaining importance in the area of medical diagnosis and to provide quality solution to solve health problems and diagnoses of various diseases. In their research work, they have utilized multilayer feed forward networks with error back propagation technique for the classification of heart disease dataset. They have trained neural network based predictive models to perform classification task of heart disease data. The adaptive technique for learning rate parameter based on adaptive momentum along with error back propagation algorithm has been applied to predict the disease. For the prediction purpose, four datasets available on UCI website such as Cleveland, Hungarian, Switzerland and long-Beach V.A datasets have been picked. According to them, testing results gives a of maximum 90.6% accuracy for single layer and 94% for multilayer feed forward network, for the classification of heart disease dataset with 13 predictor attributes and one output attribute. In their proposed work, to improve the accuracy of results they have attempted to introduce a
classification approach which includes 15 input attributes and an output attribute. They finally concluded that ANN’s ability to learn complex and non-linear relations including noisy or less precise information, neural networks are well-suited to solve problems.

2.6 Reviews on Related Works Done with ANN in Financial Sector

In recent years, researchers have explored a number of artificial intelligence techniques based on artificial neural networks to solve financial problems and these explorations appear to be promising for predictions related to financial sectors. However, further research is needed to optimize the design and information associated with ANNs based forecasting in financial sectors.

Hanifah and Faturohman have presented their research paper titled “Using artificial neural network (ANN) back propagation to predict the bankruptcy of Islamic banks in Indonesia”. In their research objectives they have focused on utilizing artificial neural networks to define the prediction of bankruptcy in Islamic Banking for the banking industry in Indonesia. This can be achieved by checking the consistency, mention the failed banks and success banks by the prediction results and by explaining the financial factors that must be improved to avoid the failures. The data used in this research has been published by the Islamic banks and the conventional banks of Indonesia. Nowadays, there has not been failure for Islamic banks in Indonesia so this research uses the data for conventional banks. Quarterly calculations of the financial ratios have been processed in the MATLAB version R2014a Neural Network Toolbox and applied neural network back propagation approach. From the results obtained, it has been observed that the average of accuracy for network model in predicting the failed bank group is 98.5% and 100% for the success bank group in the training process. From 12 banks data, the trained ANN models indicate one bank as the failed bank. Three banks have been identified in alarming situations and must pay attention to their two financial ratios. Then, there are three banks recognized and each of them must improve one ratio. Lastly, five banks have been predicted safe and successful in all financial ratios. The ANN feed-forward back propagation networks can distinguish patterns and trends based on financial ratios in making bankruptcy predictions and can be used as early warning signals to avoid failures. Performance plots and regression graphs have been generated and observed during in the training process. The Early
Warning System (EWS) based on ANN based prediction model has been successfully used to consider financial factors in making the policies or decisions in the future.

Pan et al. have used neural networks to predict the stock market successfully. They have utilized neural networks to predict the Australian stock market index (AORD – All ordinaries index) and they have attempted to design and optimize the adaptive neural networks to implement the prediction model. This effort aimed to discover an effective neural network based model for this prediction purpose, which could exploit or model various dynamical swings and inter-market influences that could be discovered from professional technical analysis and quantitative analysis. The inputs were relative returns resulting from the basic factors of the Australian stock market and inter-market influences on the Australian stock market. They were able to simulate a neural network that was able to achieve 80% accuracy for directional prediction. Authors noticed a 6-day cycle in the Australian stock market. Authors also applied artificial neural networks trained with a backpropagation algorithm to discover the optimal neural network architecture and the relative returns series of the open, high, low and closed prices in the Australian stock market. The optimal neural network architecture comprised of three layers; an input layer with 33 nodes, a hidden layer with 3 nodes and an output layer with 4 nodes. The best neural network developed in their research was able to achieve an accuracy of about 81% when predicting the next-day direction for the relative returns of open, low, high and closed prices for the Australian stock market.

Tsai and Wu have suggested using the ANN for bankruptcy prediction and mentioned about it in their research paper titled “Using neural network ensembles for bankruptcy prediction and credit scoring”. They talk about bankruptcy prediction and credit scoring by making use of neural networks based models. In their research artificial intelligence and machine learning techniques have been employed to find solution of these financial decision-making problems. The multilayer perceptron (MLP) architecture network trained by supervised backpropagation learning algorithm has been used for financial decision-making problems and is proved superior to other traditional statistical models. They also suggest combining multiple classifiers (or classifier ensembles) which can be better than single classifiers. In this paper, authors investigate the performance of a single classifier model as the baseline.
classifier to compare with multiple classifiers model and diversified multiple classifiers by employing neural networks based on three datasets. By comparing with the single classifier as the benchmark in terms of average prediction accuracy, it has been noticed that the multiple classifiers only perform better in one of the three datasets. However, for the Type I and Type II errors, there is no exact winner between the above mentioned two techniques. Authors suggest that it is better to consider these three classifier architectures or hybrid approach to make the optimal financial decision. In this paper, they have compared the performance of the single neural network classifier with the (diversified) multiple neural network based classifiers over three available datasets for the bankruptcy prediction and credit scoring problems.

Instead of predicting of stock market indexes, Khan, Bandopadhyaya and Sharma in their research work have investigated only stocks of the Tata Power Company trading on the Indian National Stock Exchange market. The authors applied and compared a back propagation neural network (BPN) and a genetic algorithm based back propagation neural network (GA-BPN) to predict the stock prices every day. The authors utilized datasets from January 2004 to 22 December 2006 for the training data set and data from January 2007 to March 2007 as testing datasets. In their research authors have claimed that the GA-BPN model has outperformed the BPN model. However, little detailed technical insight into the combination of genetic algorithms and back propagation neural networks had been provided. Required interfaces and algorithms have been implemented in Visual Basic and Microsoft Access as front end and back end tools.

Customer age and gender are critical parameters for both retailing and marketing applications. Gokhansilahtaroglu in his research paper “Predicting gender of online customer using artificial neural networks” has proposed a model to predict the gender of an online customer by analyzing his/her mouse movements. To accomplish this purpose, the author has developed a novel data cube model. The model consists of six dimensions which are customer demographic data, customer visits, mouse movements, online shopping cart, external data and time dimension. To detect customer gender he has developed and employed an artificial neural network based model. Experimental results demonstrate that using the derivatives of the data cube and the model, gender of an online customer may be predicted with up to 80% of
accuracy. In the research he has also applied a data mining decision tree analysis in order to find the most significant parameters for detecting an online customer’s gender. The analysis shows that parameters like time spent on the site, average time intervals between clicks, items clicked and order of the clicks are important and can be utilized to predict online user gender and it may be further used for promotional and marketing purposes.

An accurate prediction of stock price movement can yield attractive benefits for the investors. It usually affects a financial trader’s decision to buy or sell at right moment of time. But prediction of stock price index movement is considered as one of the challenging task of financial time series prediction. Kara, Boyacioglu and Baykan\textsuperscript{76} in their research paper “Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange” describe about prediction of stock price index movement with help of ANN. According to authors, because of the complexity of stock market data, development of efficient models for predicting is a difficult task. In their research they have attempted to develop two efficient models and also compared their performances in predicting the direction of movement in daily Istanbul Stock Exchange (ISE) National 100 Index. The models have been constructed on two important classification techniques, artificial neural networks (ANN) and support vector machines (SVM). Ten technical indicators have been selected as inputs (predictor variable) for the proposed models. Two comprehensive parameter setting experiments for both models have been performed to enhance their prediction performances. It has been emphasized that both the ANN and SVM models have shown significant performance in predicting the direction of stock price movement. Experimental results demonstrate that the average performance of ANN model (75.74\%) has been found significantly better than that of SVM model (71.52\%).

2.7 Reviews on Related Works Done with ANN in Other Sectors

Worldwide many people are now dependent on online banking. This raises a lot of security concerns as the banking sites on net are forged and fraud can be committed by identity theft. The forged sites online are termed as phishing websites and created by malicious people to mimic web pages of real websites and these attempts to defraud people of their personal information. Martin et. al.\textsuperscript{77} have mentioned in their
research paper “A Framework for Predicting Phishing Websites Using Neural Networks”, detection and identification of phishing websites is a really complex and dynamic problem involving many important factors and criterion. This research paper discusses about the prediction of phishing websites with help of neural networks. The prediction of phishing websites is essential and this can be done easily using neural networks. A neural network is a multilayer system to understand the complex relationships among phishing data, reduces the errors and enhances the system performance. This paper explains about a framework to better classify and predict the phishing sites with neural networks.

Authors have utilized neural network techniques in their new e-banking phishing website detection model to identify the most important phishing features and significant patterns of phishing characteristic or factors in the e-banking phishing website archived datasets. Each indicator ranges between the input values genuine, doubtful and legitimate. Using these values rules has been formed and the network is trained to produce output that ranges between very legitimate, legitimate, suspicious, phishy and very phishy classification.

For the prediction of phishing websites, earlier works have been done using various traditional data mining classification algorithms but the error rate of those algorithms were very high demonstrating low performances. But on the other hand, when an element of the neural networks fails, it can continue without any problem because of its parallel nature and high failure tolerance quality. In this way performance can be enhanced by considering neural networks as it reduces the error and delivers higher classification performance. Authors Aburrou, Hossain and Dahalet. al. believe that the framework works better than models based on traditional data mining techniques and gives a lower rate of error.

Kaur, Singh, Kahlonet. al. have researched on the subject of neural network based novel technique for software effort estimation. As we understand that software development is extremely important for big projects and here the authors suggest an effective and efficient method to enhance accuracy and avoid error in the development. Here the researchers have utilized soft computing based techniques and worked on NASA software data sets and also presented a comparative review between ANN, HOLSTED and DOTTY models. The results obtained in the research
depict that neural network system has the lower MMRM (Mean Magnitude of Relative Error) and RMSE (Root Mean Square Error) values in comparison to other techniques. In this paper, authors also suggest using neural based techniques to build suitable type of model and to employ this type of estimation on all types of projects.

Goswami and Bhatia in their research have focused on the comparative study of neural network models for software quality estimation. As we are aware that to achieve a high level of quality of a software is a very challenging job, so in their paper they have presented a neural network based model on object oriented metrics and the comparative study is based upon ward neural-networks (WNN), general regression neural-networks (GRNN), fuzzy neural-networks (FNN) and these models depend upon object oriented metrics. After a comparative study of the applied methods, they have found that fuzzy neural-network is giving best results as compared to other suggested models.

As mentioned by Beale, Hagan and Demuth, there is a variety of first and second order learning algorithms in artificial neural network toolbox that are available in the MATLAB as built-in functionsand these different types have different ways to update the network weights and biases. Chaipimonplin in his research paper on topic “The efficiency of using different of learning algorithms in artificial neural network model for flood forecasting at upper river ping catchment, Thailand” presents the result of exploration for the performance and efficiency of 12 different learning algorithms; Gradient Descent (GD), Bayesian Regularization (BR), BFGS Quasi-Newton (BFG), Resilient Back propagation (RP), Scaled Conjugate Gradient (SCG), Conjugate Gradient with Powell/Beale Restarts (CGB), Fletcher-Powell Conjugate Gradient (CGF), Polak-Ribiere Conjugate Gradient (CGP), One Step Secant (OSS), Variable Learning Rate Gradient Descent (GDX), Gradient Descent with Momentum (GDM), Levenberg-Marquardt Algorithm (LMA) to forecast flood at 6 and 12 hour in advances based on artificial neural network prediction Modelling. In addition, to compare the algorithmic performances, different variations for parameters like number of hidden nodes in hidden layer by 1, 50%, 75% and 100% of the number of input variables and selecting input variables with different input determination techniques; Cross correlation (C), Stepwise regression (S), Genetic algorithms (G) and combination between C and S (CS) have been included in this study. As
concluded from experimental investigation, LM and BFG have proved to be the best algorithms for flood forecasting at 6 hours rate but for 12 hour BFG is only showing the best performance with different input variables and number of hidden nodes as the maximum of $R^2$ value has been observed as 0.99 and 0.97 respectively. The crux of study is that, all 12 learning algorithms seem to achieve similar performance values but for the best performance values of learning algorithm for flood forecasting where different input variables and number of hidden nodes have no effect are the BFG and LM learning algorithms performance at 6 hrs and only BFG at 12 hrs. On the other side GD, GDM and GDX methods have shown the lowest $R^2$ values. After comparing the result with related study authors have concluded that LM and BR are the best algorithms for flood forecasting 6 and 12 hrs however BR can’t be claimed as the best algorithm because it’s performance depends number on input variables and number of hidden nodes present in the network.

Drought is a type of natural disaster which has many social and environmental impacts. It is well known that the exact time, period and intensity of drought are unpredictable, but statistical and probabilistic techniques are used for analyzing a drought in a region. Oguzturk Yildiz, and Duvan in their research paper “A drought analysis of Sivas using the standardized precipitation index (SPI) method and drought estimation with the artificial neural networks” state that forecasting of a drought in the future is a difficult task, but upcoming technologies like the Artificial Neural Networks (ANN) which can be used successfully in predicting the behavior of non-linear systems and its use has increased in the water resources engineering in recent years. In their research, drought analysis has been conducted primarily by Standard Precipitation Index (SPI) method for Sivas Province; which has a semi-arid climate. Droughts occurring in the region have been examined by calculating SPI values of monthly precipitation data of 65 years available from 1950 to 2014. Datasets have been obtained from Sivas Weather Station for determining drought characteristics (drought duration, amplitude and intensity) in different time periods of the station and their relevance to one another. In the next part of the research, forecasting of the drought in the upcoming years has been performed by the Artificial Neural Network based model. According to researchers, while there are similar types of studies in the literature, but in this research, different network models have been developed and their trials have been made and the best network model has been found as a result of these
trials. In order to check how much data can be estimated for upcoming years by the created ANN based model, data of 2007 chosen for being dry year has been used as the test data. So, when the precipitation data of the current month as input by making the necessary conversions, it has been seen that model can make predictions about the data of the next month.

Feed-Forward Back propagation Network (FFBPN) (10_6_1) model with 10 neurons in the input layer, 6 neurons in the hidden layer and 1 output neuron has been chosen to be the best network structure for single-output layered network model. In this study, three different approaches have been taken and the results found by the network have been compared. Based on the results obtained; with different configurations of network models created using the input data such as temperature, humidity, evaporation, together with precipitation data, it is concluded that more approximate estimates for the future can be made.

Singh and Saha\textsuperscript{84} have studied the role of ANN in software testing and mentioned about their study in the research paper on the topic “Application of artificial neural networks for assessing the testability of object-oriented software”. In this paper the researchers define different applications of neural networks for predicting the reliability of a software by using object-oriented metrics of design and we basically measure testability in terms of effort for testing and the software metrics include different measures like coupling, size, inheritance etc. and in this paper they have compared the predictive performance of neural networks with the two types of analysis model like least square regression and robust regression and the coding portion written in Java. They have used ANN because it takes examples of historical trends and at the end authors conclude that neural network based model has better prediction accuracy as compare to regression models.

Samborska, Alexandrov and Sieczko et al.\textsuperscript{85} in their research paper “Artificial neural networks and their application in biological and agricultural research” have shown that data analysis using artificial neural networks (ANNs) has been increasingly applied world-wide in a range of engineering and scientific fields, including agricultural and biological research. Using ANN, the analysis of results can be attained in a relatively short time, even when considering bulk volume of data. The method is evolving an attractive alternative to accepted traditional techniques and provides
accurate results that fit well with the complex pattern of prediction variables and hard-to-foretell phenomena in agricultural and biological systems. Authors further state that despite all the above mentioned opportunities, the application of ANNs in the field of agricultural and biological sciences is still limited, but it is highly expected that neural networks will evolve as of the major research tools in these fields in the coming future. Rapid developments in electronic hardware and research equipment will permit more and more researchers to process a huge amount of data, even in a short amount of time. Only ANNs will be able to deal with such bulk amounts of data to mine the trends and specific relations and behaviors present among data. In this research paper, in the literature review section author has included a variety of research material where ANNs have been applied in the field of agricultural research. Li and Yong have applied neural networks in their research on tea leaves. The authors have been able to discriminate the low quality tea leaves and to obtain a good accuracy (77.3%) classification of all three tea gardens by using ANN models that were constructed to identify tea leaf defects based on specific records. Qiao et al. have analyzed water uptake in the soil environment, by taking into consideration that water absorption by roots is reliant on the density and humidity of the soil around growing roots. Several other research studies have confirmed that the application of ANN models have been very reliable in the determination of the relative amount of water in the plant leaves. In another type of research, Zaidi et al. have used ANNs with back propagation to estimate the growth of lettuce plants. ANNs have also been used successfully for the identification of plant viruses. The results obtained indicated that the method using ANNs prove to be very reliable tool, very helpful in such types of analyses. Therefore, it has been suggested by authors to use ANN models as an alternative for traditional statistical methods used in verification of a large amount of data. Evaluation of the effects of environmental factors on banana leaves using an ANN has confirmed the usefulness of this method.

Bisi and Goyal have proposed a work in the year 2012, “Software reliability prediction using neural network with encoded input”. In this work they have proposed a software reliability model to estimate the number of failures with help of feed-forward networks. On the basis of failure count data, the execution time is encoded by using logarithmic function. In this work they tested eighteen software failure data sets and they also compare the performance with some statistical models. So we can say
that neural network has the ability to capture the non-linear patterns of the failure process by learning from the failure data and after that the data which is failed is used to learn the network to mature its own model of the failure processing with help of back propagation learning algorithm. In this paper, a feed forward neural network with two encoded schemes like exponential and logarithmic function have been proposed and after that effect of different parameters has been clarified.

2.8 Reviews on Life Insurance Sector and the Need for the Predictive Data Mining

When individuals or organizations lack sufficient knowledge or information to enable them to plan future activities, analyses of expectations or forecasting are frequently involved. When it comes to financial matters, investors must develop plans in the face of several uncertainty factors, so accuracy in forecasting is becoming a very important issue. Following paragraphs review relevant literature and explain - why forecasting in the financial areas is gaining importance especially in life insurance sector.

Guoin his research paper elaborates that data mining methodologies can improve existing actuarial insurance models by finding additional important variables, by identifying real life interactions and by detecting non-linear relationships. New data mining techniques can assist insurance firms to take crucial business decisions and turn the newly discovered knowledge into action oriented results in business practices like claim distribution analysis, asset liability management, product development, marketing and solvency analysis. According to the author some more good examples of how data mining can be useful in health insurance sector can be found in Borok’s work. According to him, data mining can be utilized mainly for the following tasks in insurance sector.

- Identification of risk factors that predict claims, profits and losses
- Customer level analysis
- Development of new product lines
- Reinsurance methodology
- Estimation of outstanding claims provision
According to the author, Data Mining procedures like ‘Deviation Detection’ and ‘Link Analysis’ can also be utilized to improve claim estimation. The estimation of the claims can be generated from a predictive model which is based on the assumptions that the future will be much like the past. Therefore historical trends and patterns present in the large historical datasets can be utilized to predict the future trends and relationships among prediction variables.

SAS reportshave shown that the impact of the financial crisis in recent decade has a dramatic effect on the insurance industry and insurance companies in turn must look at improving operational inefficiencies to combat a global soft market, plus the industry must expect aggressive regulatory activities as a result of the crisis.  

In addition, SAS reports also say that insurance is a data-rich industry; unfortunately, most of that data is underutilized. The key to gaining a competitive advantage in the insurance industry is found in analyzing this data and getting a greater insight into their business. Insurance firms can unlock the intelligence contained in their operational applications - like policy administration, claims management and CRM solutions - through modern data mining technology. Data mining uses predictive modeling, database segmentation, cluster analysis, neural networks and combinations thereof to quickly answer crucial business questions with greater accuracy. New products can be developed and marketing strategies can be implemented, enabling the insurance firm to transform a wealth of information into a wealth of predictability, stability and profits.

Sehgal, Kumar and Gupta have proposed a work on the topic “Training neural networks for insurance with conjugate gradient algorithm”, in the year 2013. Authors suggest that ANN can be successfully utilized for predictive data mining in insurance sector. In this research paper authors have explained about training of the neural network by using multilayer perceptron feed-forward neural network by using conjugate gradient algorithm, which employs supervised learning and they have use the data set from insurance sector by using MATLAB. They have used second order conjugate gradient method because this could be applied to solve non-linear and abstruse problems that are not able to solve with early and traditional methods and at the end they concluded that conjugate gradient method shows better convergence than simple gradient method.
New Data mining technology and upcoming new methods can support the insurance firms for taking crucial business decisions. Umamaheswaran and Janakiraman in their research paper “Role of Data mining in Insurance Industry” explain that insurance systems have shown tremendous and rapid developments in our society in the global era and have become as essential need for everyone. Due to the increased stress in day-to-day life, demand of various insurance sectors has increased. Data mining helps insurance firms to discovery useful patterns from the customer database. The paper aims to present - how data mining are coming useful in the insurance industry and how its techniques produce good results in insurance sector and how data mining can enhance decision making using insurance data. This conceptual paper has been written and based on secondary study, observation from various journals, magazines and reports. According to authors, Data mining plays a vital role in the entire business applications domain especially in insurance sectors. The study highlights the importance and role of data mining techniques in life insurance sector for managing the customer data and gain a business advantage. The study also explains how business decision making with data mining techniques can be implemented in the insurance domain. This paper discusses about -how insurance companies can benefit by data mining methodologies and thereby reduce overall costs, increase their profits, can acquire new customers, retain current customers and develop new policies and products. We can conclude from the study that Insurance sector is having increasing growth rate and in the stage of adopting various new data mining techniques. Use of proper data mining techniques can prove to be a boon for the insurance organizations. The future of the insurance strongly depends upon improving the service levels with usage of advanced data mining techniques.

Estimation of Insurance claims is an important and expensive problem for insurance companies. The prediction of auto insurance claims has been a challenging problem and important research area for many auto insurance companies. Weerasinghe and Wijegunasekarain their research paper titled, “A Comparative study of data mining algorithms in the prediction of auto insurance claims” try to identify the risk factors which are affected for the high number of claims because denying them may lead to increased corporate profitability and keep insurance premiums at a lower rate. The key objectives of conducting this research is to apply and investigatedifferent data mining techniques in developing a predictive model in support of auto insurance
claim prediction and also perform comparative study of the results. The research has been carried out by applying Artificial Neural Networks (ANNs), Decision Trees (DT) and Multinomial Logistic Regression (MLR) techniques to develop the prediction model for auto insurance. Comparative results indicate that the ANN is the best predictor with 61.71% overall classifier accuracy. Decision tree based model turned out to be the second best with 57.05% accuracy and the traditional logistic regression model indicated 52.39% accuracy. Optimal parameters for NN based model takes 6 input neurons and 7 minimum hidden neurons with 0.15 learning rate value. The comparative study of multiple prediction models has provided us with an insight vision into the relative prediction ability for different data mining techniques. The comparison of the results of the decision tree and neural network models has shown an interesting pattern. Policies that have been misclassified by one model have been correctly classified by the other methods. This might be an indication that the combination of the models or hybrid model may result in a better classification performance. Finally, in the study ANN based prediction model has performed the best among three models. The actual claim exposures with which the outcomes of the neural network model has been tested and compared are found satisfactory. From experimental observations, it can be concluded that the ANN based model is quite able to identify most claims with an excessive claim exposure. Though logistic regression model has shown less accuracy as compared to other methods but still it remains as good choice when the primary goal of model development is to check for possible causal relationships between predictor and predicted variables and we wish to easily understand the effect of predictor variables on the outcome.

Fraud has been seen in all types of insurance including health insurance sector. Fraud in health insurance sector is usually done by intentional deception or misrepresentation for taking some shabby advantage in the form of health expenditures. Kirlidog and Asuk, in research paper titled “A fraud detection approach with data mining in health insurance” made use of important data mining tools and techniques to detect fraud in large sets of insurance claim data. Based on a few cases that are known or suspected to be fraudulent, the data mining techniques has the ability to compute the likelihood or probability of each instance to be fraudulent by analyzing the past or previous insurance claims. The analysts can then have a closer look for such cases that have been marked doubtful by data mining
software. According to the authors, data mining techniques such as anomaly detection, clustering and classification has ability to successfully detect anomalies or outliers in large datasets. This can be very beneficial for insurance industry which has the problems with fraudulent claims. Once the anomalous claims have been marked, then several analyses can be made on them in order to conduct a thorough investigation. The main task in these predictions is to narrow down the target for detecting frauds. Such investigations can reveal some new and unknown patterns and assist insurance experts to catch defaulters. Thus we can see that fraud detection and prevention can be extremely beneficial for the insurance companies, who have to pay to the fraudsters in the form of higher insurance premiums. New data mining technologies for this task are available today and the insurance experts need to be trained for utilizing it effectively for their benefits.

2.9 Reviews and Literature on First and Second Order Gradient Based Algorithms

In recent times, field of artificial neural networks (ANNs) has attracted attention of scientists and researchers, mainly because of inventions and developments of new training techniques for neural networks, which are able to deal with large-scale and non-linear real life learning situations. Also, one more reason is their increasing popularity as intelligent techniques in variety of real life applications in all important domains.

Over the last few decades, lots of research has been done to improve the older methods and develop a variety of newer training methods; which are capable to train different types of neural networks architectures. Development of these algorithms touches a variety of subjects like optimization, numerical analysis, linear algebra, maxima minima etc. and we are also aware that there exist a number of first and second order iterative techniques for minimizing the error gradient function during training of the network. Some of the popular methods are gradient descent, Newton’s and quasi-Newton’s method, conjugate gradient techniques, scaled conjugate gradient technique, Levenberg Marquardt method etc. These techniques basically vary depending upon how they optimize the network weights and finally the learning rate parameter to minimize the error gradient and finally reach the set target value. But in
all the techniques the main aim is to reach the point of minimum error and to reach at this point in the fastest possible manner. Mathematically this requires following an iterative numerical optimization procedure to estimate optimal values for step size in each step and to search for the point of minimum error gradient.

Besides the above-mentioned techniques, adaptive learning rate based methods and adaptive momentum based methods are also commonly used training methods for artificial neural networks.

In the following paragraphs, we present review and literature for the research work done in the area of development and use of gradient based methods of first and second order and related techniques.

2.9.1 Review on Steepest Descent or Gradient Descent Method

Meza in his research paper “Steepest Descent” explains in detail about the steepest descent method which is one of the simplest and best known methods for minimizing a function. This research is very useful in understanding the basics of all gradient based approaches and new and evolving gradient based methods. This method is not very favorable for practical usage due to its slow convergence rate, but understanding the convergence behavior of this method can help to better understand many other gradient based and more sophisticated techniques.

In this paper author has given a short introduction and discussed some of the important advantages and disadvantages of this basic method. Some modifications and recent results of the steepest descent method have also been discussed. Meza discusses in the research paper about the derivation, algorithm and convergence behavior of the method. He also presents suitable examples for applying the method and about further possible extensions for steepest descent. His research paper plays an important literature in the history of gradient based techniques because all future methods basically extend the basic theory of gradient based search for the point of minimum error and on the gradient term and also try to vary the learning rate parameter in some different manner for a faster convergence.

Sehgal, Gupta and Kumar in their research paper “Minimization of Error in Training a Neural Network Using Gradient Descent Method” demonstrate the gradient descent
minimization of error energy, for training of an ANN based prediction model for insurance created on multilayer neural network with a combination of back propagation and gradient descent technique. The research explains the weight update formula and involves finding the local minimum of gradient function and thus providing a corrective adjustment of synaptic weights present among the network layers. The paper also explains the effect of learning rate parameter on convergence towards point of local minima. Learning rate parameter provides an extra control for the speed at which we fall down or descend toward the point of local minima and to regulate the step size. The application of gradient descent technique on insurance datasets finds the accuracy of the method very reliable. However, large increase of learning rate parameter shows a fail of convergence. On the other hand, if learning rate is kept too low then also it takes a long training time to descend towards the point of set target of minimum error.

2.9.2 Review on Conjugate Gradient Method and Scaled Conjugate Gradient Method

Sandhu and Chhabra in their research paper “A Comparative Analysis of Conjugate Gradient Algorithms & PSO Based Neural Network Approaches for Reusability Evaluation of Procedure Based Software Systems” have compared performances of different variations of second order conjugate gradient based methods. These variations are available as built-in algorithms in MATLAB software. In this work, particle swarm optimization (PSO) technique along with the four variants of conjugate based methods have been empirically explored to train a feed-forward neural network for estimation in reusability dataset. The performance of the trained neural networks based models for different conjugate variations have been tested to evaluate the reusability level of the procedure based software systems. The results have been recorded in terms of Accuracy, Mean Absolute Error (MAE) and Root-Mean-Squared Error (RMSE). In this paper, authors have explained in detail about algorithms their derivation and final results for different variations of conjugate gradient based techniques. Algorithms have been applied and investigated for the software reusability datasets and also the steps to create the reusability estimation models have been explained. The performance values of feed-forward neural network trained with Scaled Conjugate Gradient (SCG) method comes out to be the best with
MAE, RMSE and Accuracy values of 0.03984, 0.0484 and 84% respectively and it has outperformed the three other types of conjugate gradient variations as well as PSO trained neural network. The results obtained using the proposed system, have been found better than the results mentioned in older methods. The second best performance has been observed for Fletcher–Reeves Update Conjugate Gradient (FRUCG) algorithm with MAE, RMSE and Accuracy values of 0.0406, 0.05122 and 78% respectively. Hence, the proposed system can be successfully applied for the evaluation of reusability of the procedure based systems.

Sehgal, Gupta and Kumar in their research paper “Performance Comparison of Second Order Conjugate Algorithms in Neural Networks for Predictive Data Mining” have done a performance comparison of different variations of the non-linear conjugate gradient based methods. Prediction models for life insurance datasets have been developed in MATLAB Neural Network toolbox software and their training and performance comparison has been done with a variety of first and second order algorithms to find most efficient training algorithm, but main focus has been kept on conjugate gradient methods. Traditional second order methods like Newton’s and quasi-Newton’s method require computation of second order derivatives and need to compute hessian matrix for quadratic termination; which is very tedious and memory intense task. Here authors have applied conjugate gradient techniques; which bypass the computation of complex hessian matrix, but still achieve quadratic termination and therefore found to be efficient in terms of memory consumption.

In this research work, MLP feed forward neural network have been trained and tested with four different variations of conjugate gradient methods and have been investigated for their relative performances. The performance of Scaled-Conjugate-Gradient method (SCGM) comes out to be the best for the given datasets and it has achieved set target value of minimum gradient. Performance value for SCGM is 0.0149 and the gradient value of 9.77e-06. On the other side, it has been found from experimental investigations that first order techniques like steepest descent and its variations are not able to achieve the set target even in 1000 epochs and have shown poor performance. It has been observed that even if conjugate methods with line search couldn’t completely converge toward the set target of the order of 10^{-5}, but they have reached very near to the set target value and shown a good performance.
value. Fletcher–Reeves variation of CGM has shown second best performance and achieved a minimum gradient of 0.00255, but Powell-Beale and Polak–Ribiere updates are also pretty close. Hence, the models trained with second order conjugate methods (CGM variations) can be used effectively for the predictive data mining.

Møller in his research paper “A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning” introduced a variation of CGM which is a supervised learning algorithm with super-linear convergence rate. The algorithm is based upon second order optimization method in numerical analysis which is known as the conjugate gradient method. As an improvement over CGM, SCG utilizes second order information for error function and requires only $O(N)$ memory usage, where $N$ is the total number of weights in the network. The performance of SCG has been benchmarked against the performances of the standard error back propagation method (BP), the conjugate gradient back propagation (CGM) and quasi-Newton methods. SCG performs a speed-up of at least an order of magnitude relative to back propagation. The speed-up also depends on the convergence criterion, means the bigger is demand for reduction in error so the bigger should be speed-up. SCG avoids a time consuming line search, which CGM and BFGS methods use and consume time and memory in all iterations in order to regulate for an appropriate step size. The smaller is the complexity of the neural network set relative to the problem domain, the bigger is the possibility that the weight space will contain long ravines characterized by sharp curvature. Back propagation is not very efficient on these ravine phenomena, but it is shown that SCG handles them effectively. Tests on the 3, 4, 5, 6 and 7 bit parity problems clearly recommend that SCG scales an order of magnitude better than BP; but tests with smaller datasets and more complex problems have to be made to conclude anything certain about the scaling of SCG.

Handwritten text and character recognition has been a challenging task as compared to recognition of handwritten numeral and computer printed text because of its large variety in nature. Chel, Majumder and Nandi, have discussed in their research on “Scaled conjugate gradient algorithm in neural network based approach for handwritten text recognition” and have mentioned that neural network based approach provides most reliable performance in handwritten character and text recognition but the performance depends upon some vital factors like no of training samples, reliable
features and no of features per character, training time, variety of handwriting etc. Important features from different types of handwriting have been collected and inputted to the neural network based model for training. More number of attributes can increase testing efficiency but it may consume longer time to converge the error curve. To reduce the training time effectively proper methods should be chosen so that the system provides best training and test efficiency in lesser time that is to provide the system fastest intelligence. In this paper authors have utilized scaled conjugate gradient algorithm, a second order training algorithm for training of neural network. It has provided faster training with excellent test efficiency. A scanned handwritten text has been taken as input and character level segmentation is done. Some important and reliable features from each character have been extracted and used as input to a neural network for training.

2.9.3 Review on Levenberg Marquardet Method

Prerana et. al. in their research paper “Predictive data mining for diagnosis of thyroid disease using neural network” have presented a systematic approach for an early prediction and diagnosis of thyroid disease with help of Levenberg Marquardet and other popular training techniques along with back propagation algorithm used on MLP in neural networks. They have applied back propagation algorithm along with LevenbergMarquardt and other gradient based methods to train the neural network based models and have presented a comparative investigation for popular training methods. ANN has been trained with insurance datasets taken from a live data warehouse and testing is done with data that is not used for training process.

The gradient and performance plots have been demonstrated to show the variation of error gradient and mean square errors (MSE) verses number of epochs. From the experimental outcomes, it has been noticed that Levenberg Marquardt method has shown the best performance for achieving the set error gradient target in 59 epochs and on the other hand first order gradient decent based methods are unable to achieve the set target value of 0.0001 even in 1000epochs. Gradient descent methods have shown a poor convergence in comparison to Levenberg Marquardt method. In addition, it has also been that error accuracy limit achieved by Levenberg Marquardt method is superior as it could trains the models to accuracy level of the order of 10⁻⁵. According to authors it is also possible to extend the work in different directions like
studying the effects of varying parameters like number of layers and number of neurons in hidden layers, adaptive learning rate etc. Results show that outcomes of ANN with Levenberg technique are in good agreement with actual data; which recommends the use of ANN for earlier prediction of a disease.

In recent times, control of grinding processes in mechanical engineering by appropriate mathematical models is a key issue in metal-cutting industry. Mukherjee and Routroy in their research paper titled “Comparing the performance of neural networks developed by using Levenberg–Marquardt and Quasi-Newton with the gradient descent algorithm for modeling a multiple response grinding process” explain that monitoring and control of multiple process quality characteristics (responses) in grinding process plays an important role in precision parts manufacturing industries. Artificial neural network (ANN) based non-linear grinding process model using back propagation with gradient descent, weight adjustment algorithm (BPNN) is used widely by practitioners and researchers. However, a systematic approach to employ Levenberg–Marquardt (L–M) and Boyden, Fletcher, Goldfarb and Shanno (BFGS) update of Quasi-Newton (Q-N) algorithm for modeling of grinding process has been seldom explored. This paper explains utilization of Levenberg and BFGS methods based BPNN models for grinding process and authors have verified effectiveness of these methods by employing them in a real life industrial situation. Based on the real life datasets, the performance of Levenberg and BFGS Quasi-Newton’s methods has been compared with an adaptive learning and gradient descent method-based BPNN model. The results clearly show that Levenberg and BFGS-based networks converge faster and can predict the non-linear behavior of multiple responses grinding process with same level of accuracy as that of adaptive gradient based network.

The comparative analysis has been based on actual production data collected over a period of time. Important observations and the key findings that came out from this study are:

(i) Levenberg and Q-Newton’s method based BPNN networks are equally efficient as adaptive algorithm-based BPNN network to predict the behavior of multiple responses grinding process.
(ii) Levenberg algorithm displays the fastest convergence, followed by BFGS method and then adaptive learning algorithm.

(iii) Adaptive algorithm is of first order and relatively easy-to-understand and implementas compared to second order methods Levenberg or BFGS Q-Newton’s algorithms for online process control.

Kim in the research paper “Short-term price forecasting of Nordic power market by combination Levenberg–Marquardt and Cuckoo search algorithms” have proposed a new prediction method for short-term spot prices in the Nordic power market. Author has proposed a Cuckoo search Levenberg–Marquardt method (CSLM) for feedforward neural network (FFNN) for the solving process and combines the improved Levenberg Marquardt and Cuckoo search algorithms. The proposed model has considered actual power generation and system load as input variables to facilitate the efficient use of both transmission and power generation resources by direct market participants. During the training process, the proposed model generalizes the relationship between the area prices and the system price for the same period. The model can be further improved to track online the variation trend of the electricity price and to maintain accuracy because of the rapid training speed in CSLM training algorithm. The developed model has been tested with publicly available data acquired from the Nord Pool and the model’s performance has been compared with state-of-the-art artificial neural networks and time-series models. Besides, the proposed model has been successfully applied to forecast market-clearing price in the Spanish electricity market and to further assess the validity of the method. The results show that the new proposed CSLM-FFNN technique exhibits superior performance than other existing methods in terms of forecasting accuracy and training efficiency.

2.9.4 Review on Adaptive Gradient Based Algorithms

First order gradient based algorithms are widely used in all problem domains, but still these methods face problems of slow convergence due to fixed learning rate and don’t fit well in all situations. Rehman, Nawi and Ghazali, in their research paper “Studying the effect of adaptive momentum in improving the accuracy of gradient descent back propagation algorithm on classification problems” have suggested algorithms that make use of adaptive learning rate. From time to time, researchers
have suggested certain modifications to enhance the convergence in gradient descent backpropagation algorithm e.g. careful selection of input weights and biases, varying learning rates, momentum term, new network topologies, activation functions and ‘gain’ parameter in the activation function. In this research authors have proposed algorithm for improving the working performance of simple gradient descent with back propagation which is called ‘Gradient Descent with Adaptive Momentum (GDAM)’ by keeping the gain value fixed during all network trials. The performance of GDAM has been compared with ‘Gradient Descent with fixed Momentum (GDM)’ and ‘Gradient Descent Method with Adaptive Gain (GDM-AG)’. The learning rate has been fixed to 0.4 and maximum epochs are changed to 3000 while sigmoid activation function has been used for the experimentation. The results demonstrate that GDAM has proved a better approach than previous simple and adaptive methods with an accuracy ratio of 1.0 and tested for classification problems like WineQuality, Mushroom and Thyroid disease. Performance of GDAM has been verified by means of simulation on the three different types of classification problems like Wine Quality, Mushroom and Thyroid disease. The final results suggest that this method has fared well for all types of classification problems not only in terms of CPU consumption and epoch cycles but also has a better accuracy rate than the previous techniques.

In the research paper titled “Predicting for Sustainable Insurance with Adaptive Gradient Methods” by Sehgal, Gupta and Kumar adaptive gradient based techniques have been compared with simple first order gradient based technique and with some other important second order training techniques. Search for point of minimum error gradient, for a number of first and second order algorithms have been compared and applied to datasets taken from a live data warehouse of life insurance. Error back propagation along with adaptive algorithms has been adopted for training of multilayer feed forward networks. Comparative results have been obtained by applying these network training routines in the construction of prediction models based upon ANNs, for sustainable insurance.

The main focus of the research is to enhance the speed and accuracy of convergence of first order method. Convergence of adaptive methods has been compared with simple gradient descend and second order methods. From the results obtained it has been concluded that adaptive gradient based method prove slightly better than simple
gradient but still there performance is less and not comparable to second order techniques. Adaptive methods have taken lesser training time than simple gradient but they are not able to converge even in 1000 epochs toward error gradient of the order of $10^{-4}$. While on the other side, second order techniques are capable to meet an accuracy level of $10^{-4}$ and $10^{-5}$ and thus are better in terms of training time and accuracy. Gradient descent (GDM) with fixed learning rate has shown a poor convergence, conjugate and scaled conjugate gradient methods (CGM, SCGM) have shown the fastest convergence and performance of adaptive methods (GDA, GDM) under consideration, fall in between first and second order methods.

Saduf and Wani in their research papers “Comparative Study of High Speed Backpropagation Learning Algorithms”\textsuperscript{107} and “Improving learning efficiency by Adaptively Changing Learning Rate and Momentum”\textsuperscript{108} explain that Error Backpropagation which is one of the well-known training algorithms for artificial neural networks, but rate of convergence in back propagation learning is relatively slow, which in turn makes this method computationally excruciating. Over the last few years a number of modifications have been suggested by researchers to enhance the efficiency and convergence speed of the back propagation algorithm.

The main focus of this paper is on investigating and comparing the performances of improved versions of back propagation algorithm in training the neural network. All of them are assessed on different training datasets and a comparative investigation has been made. Results of computer simulations with standard benchmark problems like XOR, 3 BIT PARITY, MODIFIED XOR and IRIS have been tested and presented. The training performances of these improved methods have been evaluated in terms of percentage of accuracy and convergence speed. Learning rate parameter is the most important factor and plays crucial rule during convergence of the method for set target. In the modified algorithms learning rate has been made adaptive in a number of ways and performances of these variations have been compared. In this research, authors have implemented back propagation algorithm with adaptive learning rate and adaptive momentum techniques and tested the new improved methods on number of training sets to determine its performance in terms of convergence speed and accuracy. Implementation results demonstrate that backpropagation with adaptive
Learning rate and adaptive momentum outperforms the simple backpropagation with simple gradient descent which uses a fixed learning rate.

Exact diagnosis and classification of breast tumors is very important to reduce inadequate surgeries and unnecessary number of biopsies. Gradient descent based algorithms in artificial neural network has been successively employed for classification of breast tumors by a number of researchers. Singh, Verma and Thoke have presented use of adaptive methods in their research article “Adaptive gradient descent back propagation for classification of breast tumors in ultrasound imaging”. In this paper comparative evaluation for the performance of three gradient descent based artificial neural network (ANN) based techniques to classify the tumor as benign and malignant in ultrasound imaging has been presented. The ultrasound images have been preprocessed by wavelet filters for reducing the speckle noise. Fifty seven texture and shape attributes have been extracted from filtered breast ultrasound images to classify the breast tumors. Classification accuracy, sensitivity and CPU times have been used as figure of merit for the classifier. Observations and results indicate that adaptive gradient descent back propagation based on variable learning rate has outperformed other techniques while giving highest classification accuracy of 84.6%. However appropriate choices for parameters like learning rate, network model and architecture for a given classification task are still an important issue. This work has presented a comparative analysis of gradient descent based neural network techniques for classifying breast tumors using 40 texture and shape features. Experiments have been conducted on new database of 89 historically confirmed real breast ultrasound images containing 44 benign and 45 malignant breast cases and using MATLAB Neural Network Toolbox platform. Three gradient based algorithms namely gradient descent (GD), adaptive gradient descent (AGD) and gradient descent with momentum (GDM) have been evaluated in terms of AUC, sensitivity, specificity, accuracy and CPU time. It has been shown that gradient descent back propagation neural network can be successfully employed for classification of breast tumors. It has been found that AGD based back propagation has outperformed other two methods due to its highest capability in classifying benign and malignant tumors and has achieved the classification accuracy up to 84.6%. But on the negative side, this method suffers from time complexity, hence in future we may have to think to evaluate some other classification algorithms which can converge faster and at the
same time can give high classification accuracy. Further the performance of classifiers on bulky dataset with more number of descriptive features must be evaluated.

Hamid et. al.\textsuperscript{110} in their research article “Accelerating learning performance of back propagation algorithm by using adaptive gain together with adaptive momentum and adaptive learning rate on classification problems” present a number of methods to accelerate the search speed for reaching toward point of minimum of gradient. These methods include considering a number of parameters like effect of gain parameter on the performance and effect of momentum parameter and adaptive learning etc. They have also presented stepwise algorithms in this chapter. They have also applied the suggested techniques on a number of standard problems. They have mentioned algorithms like The Back Propagation Gradient Descent (BPGD), The Back Propagation Gradient Descent with Adaptive Gain (BPGD-AG), The Back Propagation Gradient Descent with Adaptive Gain, AdaptiveMomentum and Adaptive Learning Rate (BPGD-AGAMAL). For comparing the performances of the proposed algorithms with conventional BPGD and BPGD-AG methods, the network parameters such as network size and architecture (number of nodes, hidden layers, etc.), values for the initial weights and gain parameters have been kept the same. Experimental results display that the BPGD-AGAMAL perform better as compared to BPGD and BPGD-AG. Moreover, when comparing the proposed algorithm with BPGD and BPGD-AG, it has also been empirically shown that the proposed algorithm (BPGD-AGAMAL) outperformed with highest accuracy than BPGD and BPGD-AG methods. This conclusion recommends the usage of the suggested algorithm as an alternative training algorithm for neural networks based models.

It may be seen from these studies that this area of research is yet to be explored extensively and we have the need to focus on the development of new methods for predictive data mining and search for better training algorithms for neural networks based predictive modeling. However, to the best of my knowledge & belief, none of the earlier research has been made on the research topic under consideration and the present research work will try to fill up this gap.