Classification of Road Accident Patterns

Anuj Sharma[1] Dr. Sandeep Sharma[2] Dr. Santosh Kumar[3]
[1,2,3]Department of Electronics & Communication Engineering
[1,2,3]DIT University Dehradun, Uttarakhand, India

Abstract— This paper presents the classification of vehicular road accident patterns using data mining in Uttarakhand state. In order to provide insights for the development of safety improvement strategies. At this aim, data mining techniques are used to analyze the data relative to the 4,130 crashes all including vehicular crashes. Understanding the accidents on age based classification under which drivers and passengers are more likely to be killed or more severely injured in an automobile accident can help improve the overall driving safety situation. Data mining a non-trivial extraction of novel, implicit, and actionable knowledge from large data sets is an evolving technology which is a direct result of the increasing use of computer databases in order to store and retrieve information effectively. It also enables data exploration, data analysis, and data visualization of huge databases at a high level of abstraction, without a specific hypothesis in mind. Understanding the casualties and classifying the road accidents on gender basis can help in improving traffic safety conditions. This paper presents a data mining approaches to explore various factors for vehicular collision. We have used neural network to classify various road accident patterns that can in turn help in analyzing the road accidents and casualties according to age, gender and location within the state. Feed forward neural network is used to evaluate the various results regarding the related study. With the use of data mining techniques it would be analyze and observe the large database having different records of vehicular road crashes in Uttarakhand state.

Key words: Road Accident, Data Mining, Neural Network

I. INTRODUCTION

The ever increasing tremendous amount of data, collected and stored in large and numerous data bases, has far exceeded human ability for comprehension without the use of powerful tools. Consequently, important decisions are often made based not on the information rich data stored in databases but rather on a decision maker's intuitions due to the lack of tools to extract the valuable knowledge embedded in the vast amounts of data. This is why data mining has received great attention in recent years. Data mining involves an integration of techniques from multiple disciplines such as database technology, statistics, machine learning, high-performance computing, pattern recognition, neural networks, data visualization, information retrieval, image and signal processing, and spatial data analysis [1].

Road safety experts and researchers deal with large volumes of quantitative information and collected statistics, in order to understand and estimate the social and economic cost of the accidents and to be able to introduce safety plans in order to prevent or reduce occurrences of accidents. The road traffic and accident statistics must be presented in such a way to make it easier to be both recognized and interpreted by a human operator. Previous works on accident analysis included statistical methods and formal techniques. Statistics tables and ordinary charting techniques are not sufficient for present day requirements and this causes difficulties in the effective visualization of results and patterns. Another disadvantage is that ordinary methods limit human involvement in the exploration tasks [2].Traffic control system is one of the various areas, where critical data about the well-being of the society is recorded and kept. Various aspects of a traffic system like vehicle accidents, traffic volumes and concentration are recorded at different levels. In connection to this, injury severities resulted from road traffic accident are one of the areas of concern. The costs of fatalities and injuries due to traffic accidents have a great impact on the society. In recent years, researchers have paid increasing attention to determining factors that significantly affect severity of driver injuries caused by traffic accidents [3]. In the paper Neural Network is used to evaluate results and perform mining.

II. RELATED WORK

An algorithm was proposed to study the application of BNN models for predicting motor vehicle crashes. To accomplish this objective, a series of models was estimated using data collected on rural frontage roads in Texas. Three types of models were compared: BPNN, BNN and the Negative Binomial (NB) regression models. The results of this study presented that in general both types of neural network models perform better than the NB regression model in terms of data prediction. Although the BPNN model can occasionally provide better or approximately equivalent prediction performance compared to the BNN model, in most cases its prediction performance is worse than the BNN model. In addition, the data fitting performance of the BPNN model was consistently worse than the BNN model, which suggested that the BNN model had better generalization abilities than the BPNN model and could effectively alleviate the over-fitting problem without significantly compromising the nonlinear approximation ability. The results also proposed that BNNs could be used for other useful analyses in highway safety, including the development of accident modification factors and for improving the prediction. The paper first described the fundamental principles of NB regression models commonly used in highway safety and the characteristics of neural network models. The review has shown that although neural network models have excellent function approximation abilities and do not require specifying a functional form linking the dependent variable to the explanatory variables, the over-fitting problem has significantly limited their application in highway safety [4]. Another paper presented the analysis of powered two-wheeler (PTW) crashes in Italy in order to detect interdependence as well as

dissimilarities among crash characteristics and provide insights for the development of safety improvement strategies. Data mining techniques were used to analyze the data relative to the 254,575 crashes involving PTWs occurred in Italy in the period 2006–2008. Classification trees analysis and rules discovery were performed. Tree-based methods are non-linear and nonparametric data mining tools for supervised classification and regression problems. They do not require a priori probabilistic knowledge about the phenomena under studying and consider conditional interactions among input data. Rules discovery is the identification of sets of items that occur together in a given event more often than they would if they were independent of each other. Both the classification trees and the rules discovery were effective in providing meaningful insights about PTW crash characteristics and their interdependencies. Even though in several cases different crash characteristics were highlighted [5]. Dursun Delen, Ramesh Sharda, Max Bessonov [4] proposed factors that affect the risk of increased injury of occupants in the event of an automotive accident include demographic or behavioral characteristics of the person, environmental factors and roadway conditions at the time of the accident occurrence, technical characteristics of the vehicle itself, among others. This study used a series of artificial neural networks to model the potentially nonlinear relationships between the injury severity levels and crash-related factors. It then conducted sensitivity analysis on the trained neural network models to identify the prioritized importance of crash-related factors as they apply to different injury severity levels. In the process, the problem of five-class prediction is decomposed into a set of binary prediction models (using a nationally representative sample of 30,358 police-recorded crash reports) in order to obtain the granularity of information needed to identify the "true" cause and effect relationships between the crash-related factors and different levels of injury severity. Eight binary MLP neural network models were developed with different levels of injury severity as the dependent variable. These eight models presented different levels of injury severity varying from the no-injury to fatality and from fatality to no injury. All models were found to have better predictive power as compared to a model with a five-category outcome variable. In addition, this structure helped to identify the important explanatory variables at each level of distinction between the injury severities. Another study intended to provide insight into pedestrian accidents by uncovering their patterns in order to design preventive measures and to allocate resources for identified problems. Kohonen neural networks are applied to a database of pedestrian fatal accidents occurred during the four-year period between 2003 and 2006. Results showed the existence of five pedestrian accident patterns: (i) elderly pedestrians crossing on crosswalks mostly far from intersections in metropolitan areas; (ii) pedestrians crossing suddenly or from hidden places and colliding with two-wheel vehicles on urban road sections; (iii) male pedestrians crossing at night and being hit by four-wheel vehicles on rural road sections; (iv) young male pedestrians crossing at night wide road sections in both urban and rural areas; (v) children and teenagers crossing road sections in small rural communities. From the perspective of preventive measures, results suggested the necessity of designing education and information campaigns for road users as well as allocating resources for infrastructural interventions and law enforcement in order to address the identified major problems [6]. DTs allow accident classification based on crash severity. They provide an alternative to parametric models due to their ability to identify patterns based on data, without the need to establish a functional relationship between variables. Moreover, such classification models can be used to determine interactions between variables that would be impossible to establish directly, using ordinary statistical modelling techniques [7]. Miao M. Chong, Ajith Abraham, Marcin Paprzycki [8] presented in a paper the severity of injury resulting from traffic accidents using artificial neural networks and decision trees. They applied them to an actual data set obtained from the National Automotive Sampling System (NASS) General Estimates System (GES). The experiments also showed that the model for fatal and non-fatal injury performed better than other classes. The ability of predicting fatal and non-fatal injury is very important since drivers' fatality has the highest cost to society economically and socially. It has to be stressed again that it is a well-known fact that one of the most important factor differentiating injury level is the actual speed that the vehicle was going when the accident happened. Our dataset doesn't provide enough information on the actual speed, since speed for 67.68% of the data records was unknown. If the speed was available, it might have helped to improve the performance of the two considered models. In another study [9] crash data was mined to identify the driver and vehicle attributes which are the main causes for road accidents. Principal Component Analysis was used to emphasize the relationships between characteristics such as age, gender and vehicle type, to the crash variables.

III. METHODOLOGY

A. Artificial Neural Network

ANN can be defined as a system designed to model the method to fulfill as a function of brain. An ANN is formed by being connected in various shapes of artificial neural cells with each other. ANN has an ability of gathering information, saving and generalization this information with connection weights between the cells after passing learning algorithm and learning process. ANN method is generally used for modeling the variables which are in nonlinear relation and good results are obtained. Since the variables used in this study are in nonlinear relation, ANN method is preferred. In the application of ANN, a program code was written by using MATLAB software. In the program, software was developed being able to do cycling among transferring, training functions and in the number of neuron (1, 2, 3, 4…..) in hidden layer. In the structures of alternative network input layer and neuron number in hidden layer, output layer, training function, transferring functions among the layers, R2, and MSE values were provided with its being able to be read in Microsoft Excel[12]. Features and superior sides of ANN are being nonlinear, parallelism, easiness of being reality, processing of local information, error tolerance, generalization, adaptation, hardware speed, learning, analysis and easiness of design (Therefore, the most suitable network structure can be easily determined).

Artificial neuron performs the following:

- Receives signal from other neurons.
- Multiplies each signal by the corresponding connection strength, which is weight.

- Sums up the weighted signals and pass them through an activation function.
- Feeds output to other neurons.

Feed forward networks feed outputs from individual neurons forward to one or more neurons or layers in the network. The output of a node is scaled by the connecting weight and is fed forward as an input to the nodes in the next layer of the network. The input layer plays no computational role but merely serves to pass the input vector to the network. The input layer and the hidden layer are connected by weights and likewise the hidden layer and output layer also have connection weights. The network has the ability to learn through training. The training requires a series of input and associated output vectors. During the training, the network is repeatedly presented with the training data and the weights and thresholds in the network are adjusted from time to time till the desired input output mapping occurs.

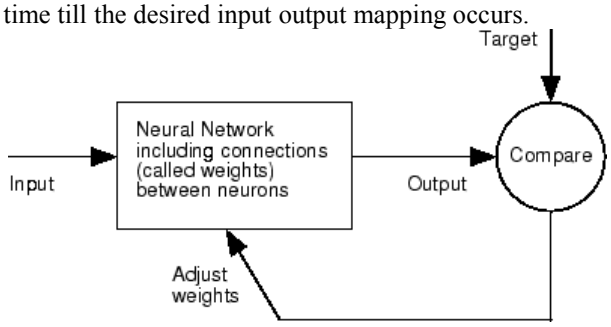

Fig. 1: Artificial neural network

Typically, neural networks are adjusted, or trained, so that a particular input leads to a specific target output. The next figure illustrates such a situation. There, the network is adjusted, based on a comparison of the output and the target, until the network output matches the target. Typically, many such input/target pairs are needed to train a network is chosen using a variety of impurity or diversity measures.

IV. DATA

The study data includes the 4,130 records of the victim of Uttarakhand state including fields like name, age, gender, location, RTA type, region, district, and block. The data is used to analyze the accident patterns on age and gender basis. Also including the areas which are more affected in the state of Uttarakhand.

V. RESULTS AND DISCUSSIONS

The neural network is used to analyze the data and perform data mining. We have shown 4 different models for male and female respectively, varying the number of neurons. Models with neuron 8, 14, 16 and 20 have been trained and compared to analyze the accident data according to age and gender.

Further analysis is done on the basis of age groups as follows:

- 0 – 15 years as group 0
- more than 15 - 35 years as group 1,
- more than 35 - 50 years as group 2 and
- More than 50 years as group 3.

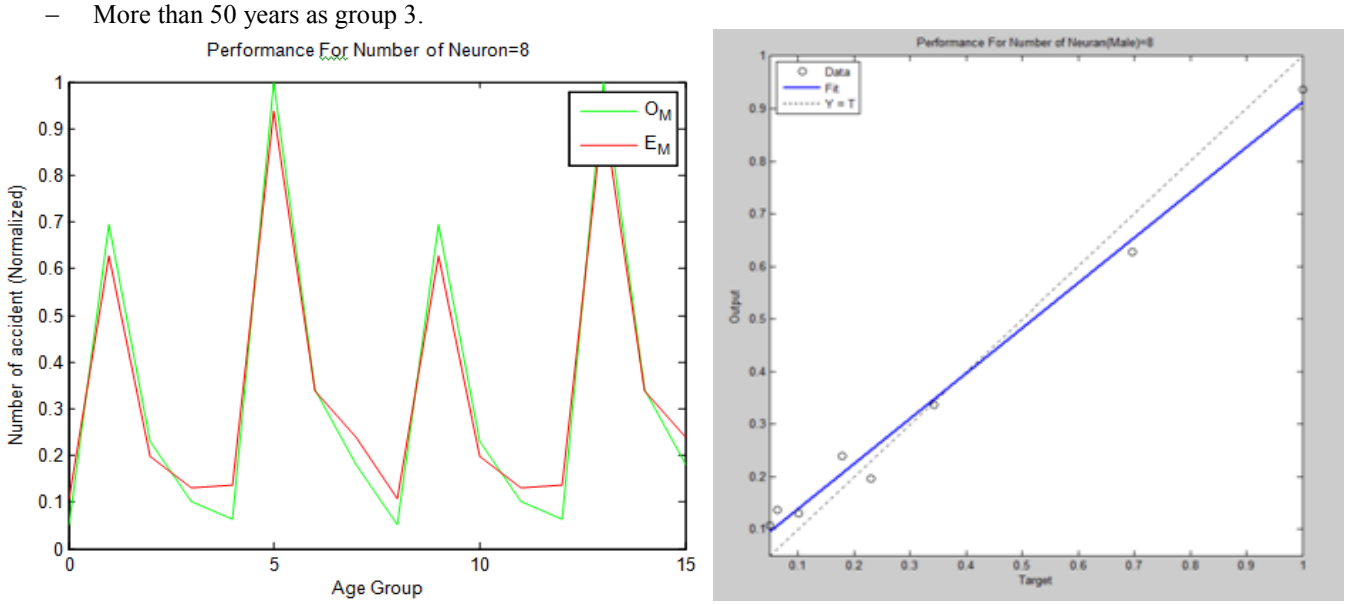

Fig. 2: Model with Neuron (male) = 8

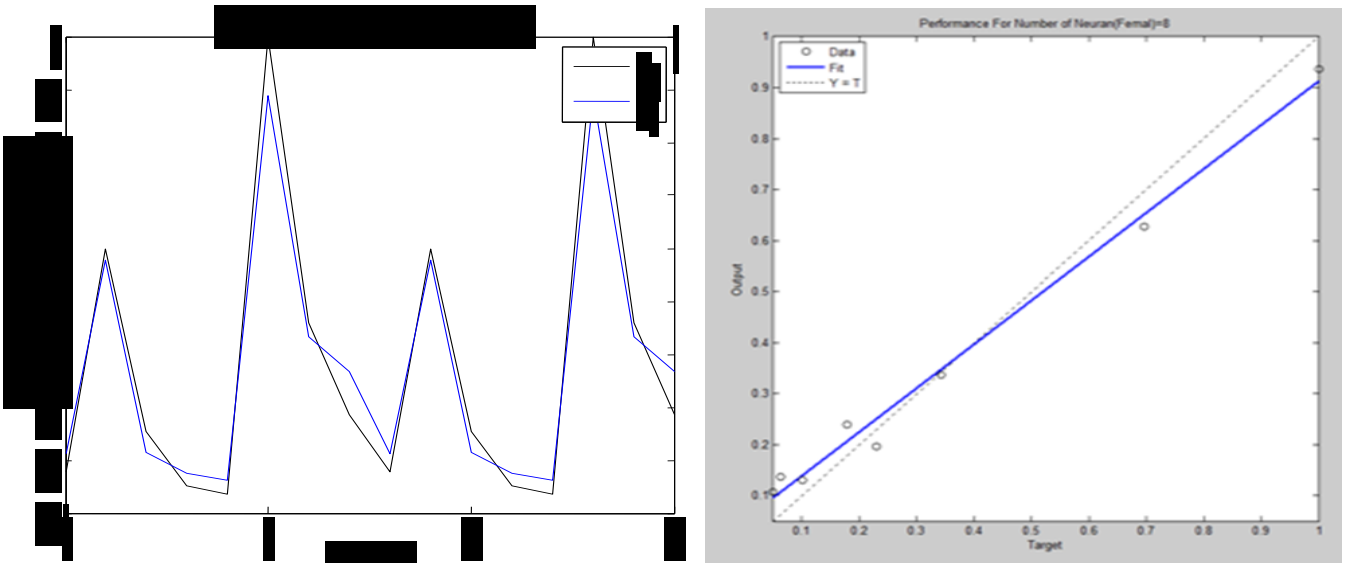

Fig. 3: Model with Neuron (female) =8

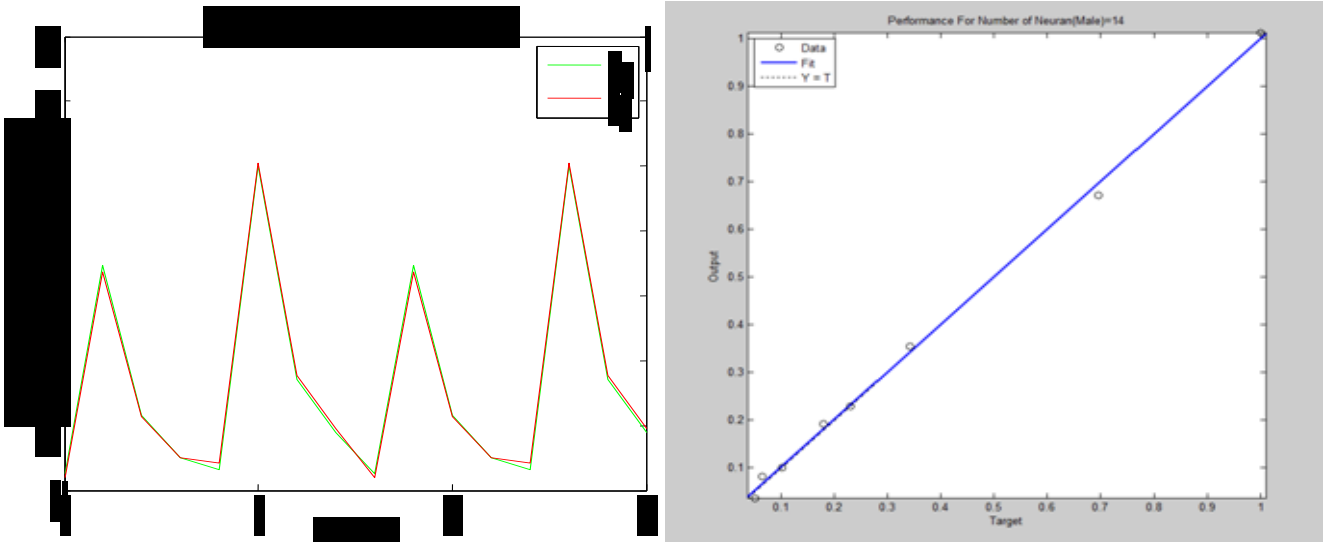

Fig. 4: Model with Neuron (male) = 14

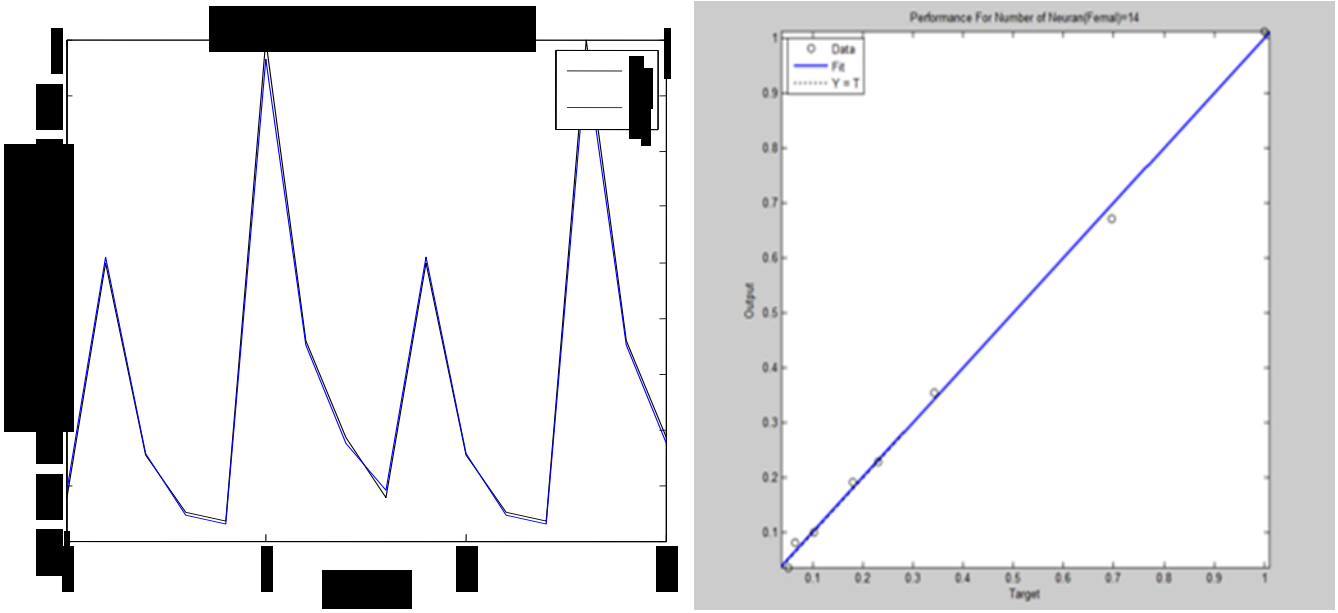

Fig. 5: Model with Neuron (female) =14

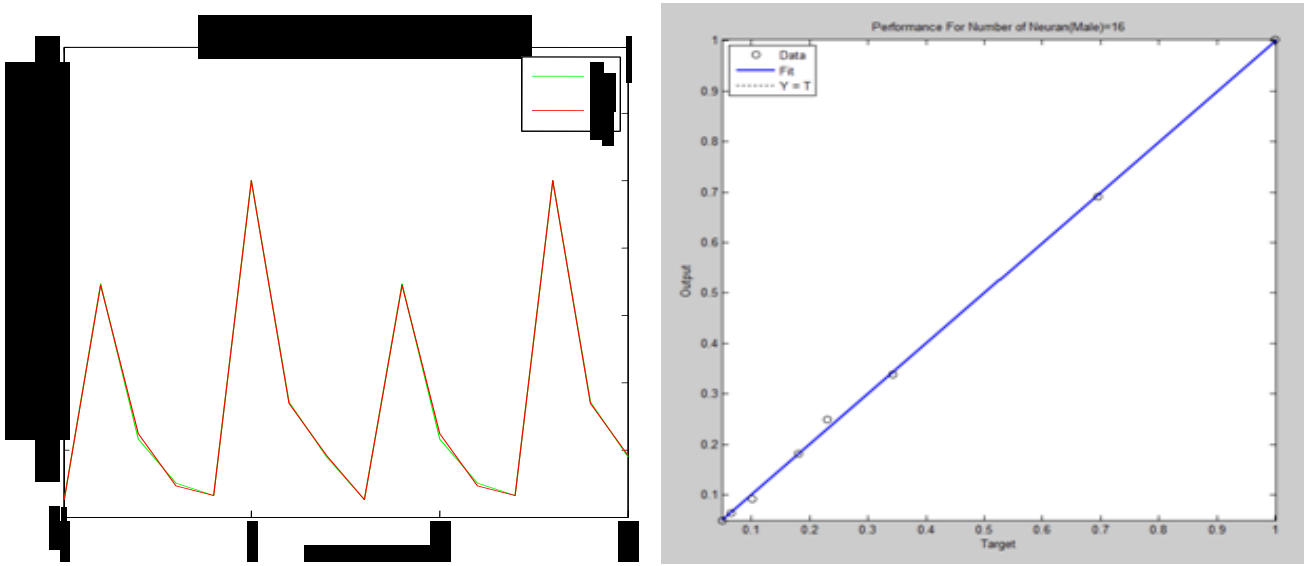

Fig. 6: Model with Neuron (male) = 16

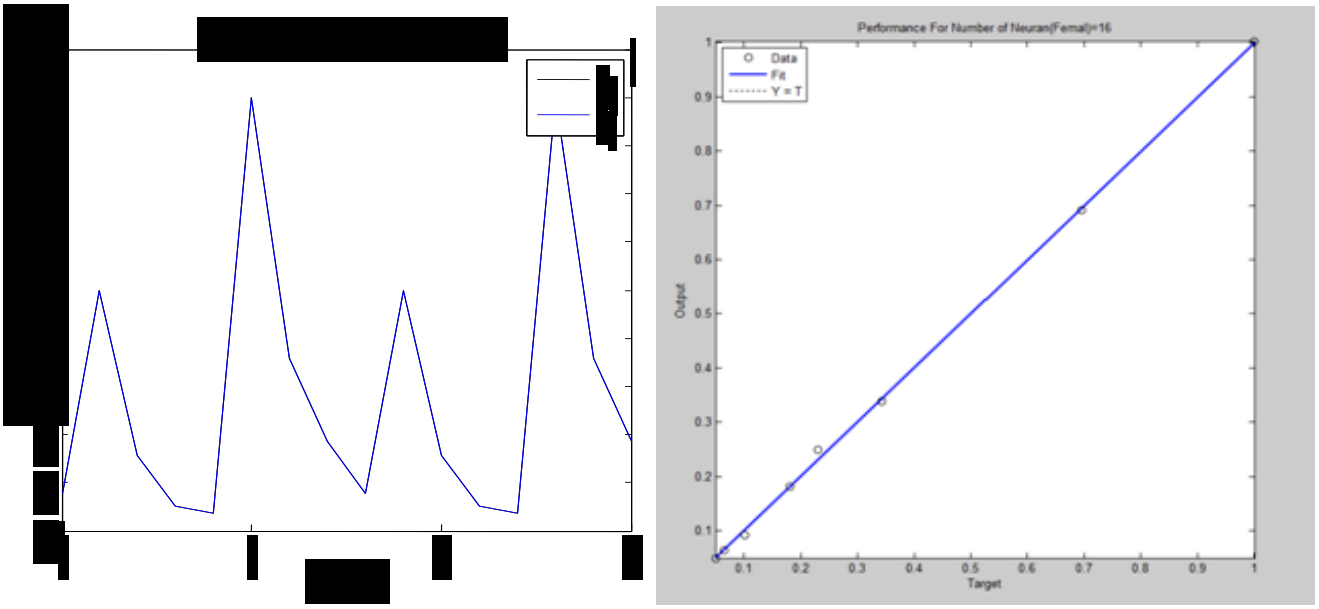

Fig. 7: Model with Neuron (female) = 16

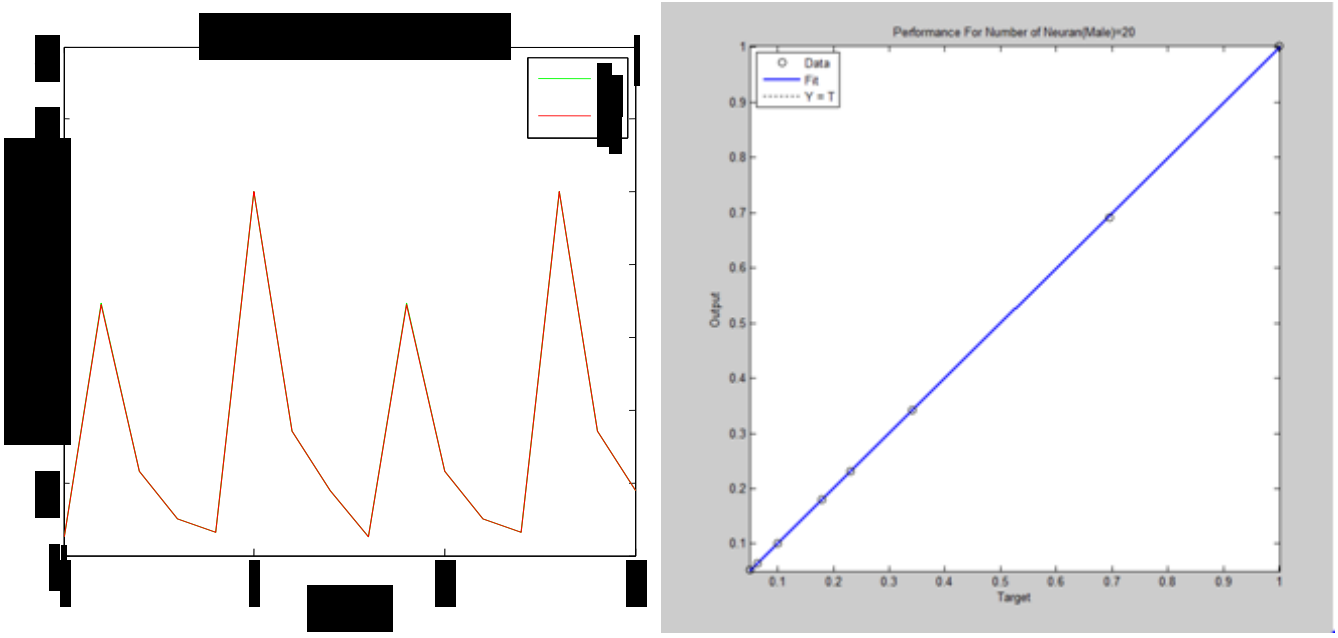

Fig. 8: Model with Neuron (male) =20

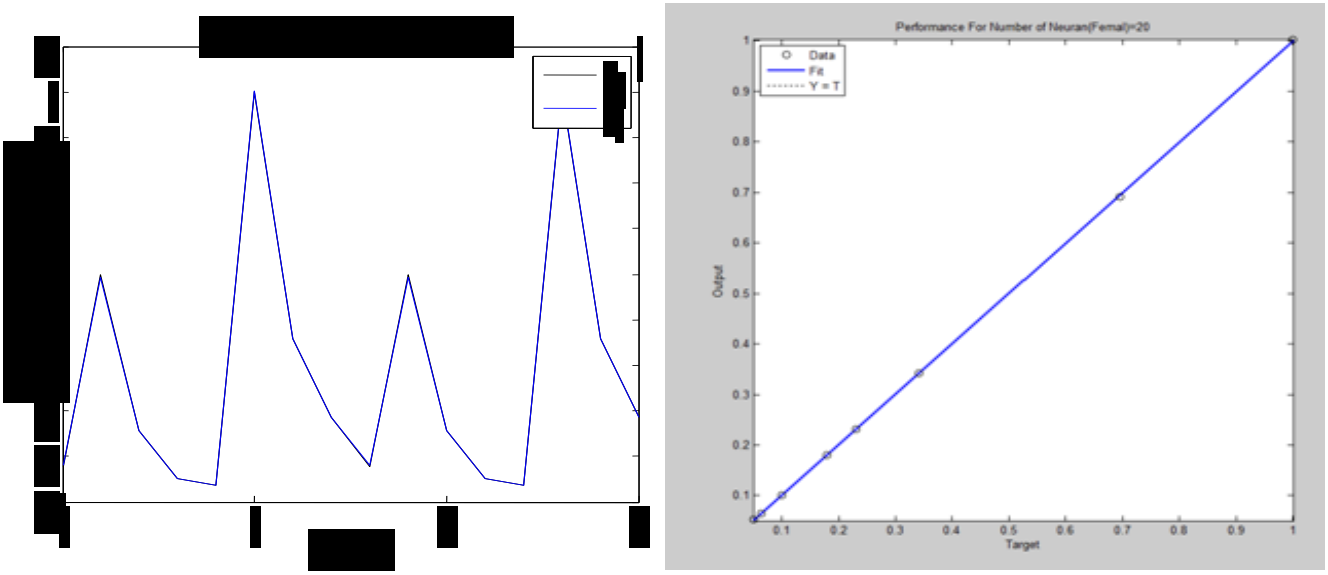

Fig. 9: Model with Neuron (female) = 20

VI. CONCLUSION

In this paper feed forward neural network algorithms is implemented. Above we have shown the different models by varying the number of neurons in male and females differently to classify the accident patterns in Uttarakhand. The model with neuron 20 is the best model according to the accuracy and performance. Further it gave nearest to accurate results when compared to target results. Based on the study it is observed that number of males affected is much more than number of females affected in the accident. The study has also analyzed that most number of cases have occurred in Garhwal district followed by Kumaon region. Also the study shows that age group of 15 years to 35 years is most affected by accidents.

VII. FUTURE SCOPE

Neural network is very less implemented for data mining. The research can further continue using different neural algorithms with varying training functions and iterations, to lesser the complexity and increase performance and efficiency.

REFERENCES

[1] S. Shanthi, G. Ramani, "Classification of Vehicle Collision Patterns in Road Accidents using Data Mining Algorithms", International Journal of Computer Applications (0975 – 8887) Volume 35– No.12, December 2011.
[2] K. Jayasudha and C. Chandrasekar, "An Overview of data mining in road traffic and accident analysis", Journal of Computer Applications, Vol – 2, No.4, Oct – Dec 2009.
[3] A. Montella, M. Aria, A. D'Ambrosio, F.Mauriello, "Analysis of powered two-wheeler crashes in Italy by classification trees and rules discovery", Accident Analysis and Prevention 49 (2012) 58– 72.
[4] D. Delen, R. Sharda, M. Bessonov, "Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks", Accident Analysis and Prevention 38 (2006) 434–444.
[5] C.G. Prato, V. Gitelman, S. Bekhor, "Mapping patterns of pedestrian fatal accidents in Israel", Accident Analysis and Prevention 44 (2012) 56– 62.
[6] Miao M. Chong, Ajith Abraham, Marcin Paprzycki, "Traffic Accident analysis using decision trees and neural network."
[7] Kim, Nitz, J. Richardson, & L. Li, "Personal and behavioral Predictors of Automobile Crash and Injury Severity" Accident Analysis and Prevention, Vol. 27, No. 4, 1995, pp. 469-481.
[8] Shankar, Mannering , & Barfield, "Statistical Analysis of Accident Severity on Rural Freeways" Accident Analysis and Prevention, Vol. 28, No. 3, 1996, pp.391-401.
[9] S.Y. Sohn, S.H. Lee, 2003, "Data fusion, ensemble and clustering to improve the classification accuracy for the Severity of road traffic accidents in Korea". Safety Science, 4(1), pp. 1-14.
[10] S. Soni, "implementation of multivariate data set by cart", International Journal of Information Technology and Knowledge Management. July-December 2010, Volume 2, No. 2, pp. 455-459.
[11] J.R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, San Mateo, CA, 1993.
[12] Halim Ferit Bayata, Fatih Hattatoglu and Neslihan Karsli, "Modeling of monthly traffic accidents with the artificial neural network method", International Journal of the Physical Sciences ,Vol. 6(2), pp. 244-254, 18 January, 2011