

COMPARATIVE STUDY BETWEEN DECISION TREES AND NEURAL NETWORKS TO PREDICT FATAL ROAD ACCIDENTS IN LEBANON

Zeinab FARHAT¹, Ali KAROUNI², Bassam DAYA³ and Pierre CHAUVET⁴

¹EDST, Lebanese University, Lebanon, Beirut

²University Institute of Technology, Lebanese University, Lebanon, Sidon

³University Institute of Technology, Lebanese University, Lebanon, Sidon

⁴LARIS EA, Angers University France, France, Angers

ABSTRACT

Nowadays, road traffic accidents are one of the leading causes of deaths in this world. It is a complex phenomenon leaving a significant negative impact on human's life and properties. Classification techniques of data mining are found efficient to deal with such phenomena. After collecting data from Lebanese Internal Security Forces, data are split into training and testing sets using 10-fold cross validation. This paper aims to apply two different algorithms of Decision Trees C4.5 and CART, and various Artificial Neural Networks (MLP) in order to predict the fatality of road accidents in Lebanon. Afterwards, a comparative study is made to find the best performing algorithm. The results have shown that MLP with 2 hidden layers and 42 neurons in each layer is the best algorithm with accuracy rate of prediction (94.6%) and area under curve (AUC 95.71%).

KEYWORDS

Data mining, Fatal Road Accident Prediction, Neural Networks, Decision trees.

1. INTRODUCTION AND RELATED WORK

The most frequent fatalities in worldwide are recorded on roadway accidents, according to World Health Organization report stating that more than 1.25 million of deaths and more than 20-50 million of injuries are reported every year [1]. In Lebanon, about 1100 people died and 9914 injured in road accidents in 2016 and 2017 based on the data set collected by Lebanese Internal Security Forces (ISF). These outcomes have led to large negative economic and social impacts. The resolving of this damage imposes reducing the frequency of road accidents by intensification of researches about the main factors influencing the occurrence of fatal accidents. In recent years, several data mining techniques were applied on big data sets to analyze, classify and to predict the important interactive causes of road accidents. Growth of data information was an opportunity to engage on the process of hidden knowledge extraction from a large data, whereas recognized by data mining. Data mining is known as knowledge discovering from data (KDD). It is advancement in information technology applications to analyze amount of data, make a decision and to improve precision in various fields' specialized public affairs such as finance, marketing,

science, transportation, etc.... Road accident is one of world's significant public phenomena. Shanthi and Geetha Ramani [2] worked on a comparative study of various classification algorithms such as CR-T, ID3, CS-CRT, CS-MC4, Naïve Bayes classifier and Random Tree according to the misclassification rate with and without Feature Ranking algorithm and with Arc-X4 Meta classifier. This study was applied on road accident data set provided by Critical Analysis Reporting Environment (CARE) system in which the period 2005- 2009 consists of 457549 samples and 33 attributes for 56 states in U.S.. The results showed that Random Tree classifier with Feature Ranking and Arc-X4 Meta classifier giving 0.27% error rate (accuracy 99.73%) outperformed all other algorithms and identified the main causal factors to reduce fatality rate: collision, seating position, harmful event, protection system, age range and drug involvement. In order to analyze, classify and cluster the states/regions in U.S. according to high and low fatal rate of accidents, Liling et.al [3] applied association rule mining, apriori algorithm, and Naïve Bayesian classifier after several statistics from the data set provided by FARS (Fatal Accidents Reporting System). They highlighted five principal attributes (collision type, speed limit, weather, surface condition, light condition and drunk driver) which represented a higher fatal rate. The analysis proved that driver drunk (yes/no) and collision type is more effective on fatality rate. In addition, the clustering results by using simple K- means algorithm showed a higher fatal rate in the southern region compared to the rest regions respectively west, Midwest and northeast. A perspective analysis of traffic accident took place by Krishnaveni and Hemalatha [4] to predict the severity of accidents by comparing the results of some classification models: Naïve Bayes, J48, AdaBoost M1 Meta, PART and Random Forest Tree. These techniques were applied on 34575 samples of road accidents in Hong Kong for the year 2008 that are categorized into three categories respectively accident, vehicle and casualty. This perspective showed that Random Forest Tree outperforms the other models of classification based on the higher accuracy rate of correct classification. Hamdeh et.al [5] was made a comparative study of performance of several data mining applications such as Neural network (MLP) in one hidden layer, Decision tree J48, Fuzzy Logic, Linear Discriminant Analysis (LDA), and Support Vector Machine (SVM). This study demonstrated that decision tree outperformed all others with a high accuracy (97.8%) in forest fire prediction as aim to decrease the fire occurrence in Lebanon by using four main meteorological parameters (Temperature, Humidity, Precipitation, and Wind speed).

Additionally, data mining techniques seek to solve complex issues that comprise a huge data set in various domains [6][7]. Road accidents are one of these complex phenomena that look to identify the main influencing factors and to predict the severity of the injury. Bahiru et al. [8] studied in their research the performance of some classification algorithms: Decision Trees (ID3, J48, CART) and Naïve Bayesian classifier according to the accuracy rate for three values of the class attribute "Accident Severity" (Fatal, Serious and Slight). Whereas J48 overpasses other classifiers in term of accuracy but Naïve classifier has better AUC than other algorithms. Moreover, the analysis of results showed that speed limit, weather condition, number of lanes, lighting condition and accident time are factors of effectiveness in road accidents. On other side, in order to build a suggestion about the influencing factors of road accidents in Saudi Arabia particularly in Riyadh city, Al-Turaiki et al. [9] implemented three classification models: CHAID, J48 and Naïve Bayes. They evaluated their performance based on accuracy rate. The results revealed that J48 (accuracy 98.26%) exceeded other models and distraction was found the most important factor for fatal accidents in Riyadh. The most important measures for the performance evaluation of machine learning algorithm are accuracy rate, error rate, precision, recall, mean absolute error (MAE) and root mean squared error (RMSE). Suganya and Vijayarani [10] used these measures in their comparative study for several classification models respectively named linear regression, logistic regression, decision tree, SVM, Naïve Bayes, KNN, Random Forest and gradient boosting algorithm. These algorithms were applied on road accident data in India from the year 2011 to 2016 by using R tool. This study revealed that KNN algorithm is the best enrolling high accuracy 93.7% and low MAE 0.0024. Baluni and Raiwani [11] implemented artificial neural network to analyze 4000 samples of accidents in Uttarakhand (India) on the basis

of age and gender attributes. A feed forward neural network was developed with two different training functions by using MATLAB software. The evaluation of these models based on each gender showed that the number of males in the accidents exceeds numbers of females. The research analysis of Olutayo and Eludire [12] revealed that decision tree ID3 surpassed all others models (artificial neural network (RBF and MLP) and decision tree (FT)) and the most important causes of accidents in one of Nigeria's most crowded roads are tyre burst, loss of control and over speeding. An accident prediction model in Erzurum's highways in Turkey was put under study by Çodur and Tortum [13] using a data set collected for the years 2005-2012 by Directory of Erzurum Traffic Region. These researchers used five different models of artificial neural network with sigmoid activation function and non-linear output function. They then evaluated the performance of each model by analyzing certain criteria of correlation, R-squared, MSE, and RMSE. The results manifested that the degree of vertical curvature of the highway is the main factor contributing to accidents on the highway whereas the degree of horizontal curvature comes next in the orderliness.

This paper divided into two sections as follows : In the first section we applied two different algorithms of Decision Trees (C4.5 and CART) and Artificial Neural Network multilayer perceptron MLP with various numbers of neurons in one hidden layer and two hidden layers to predict fatality of road accidents in Lebanon. In section 2 a comparative study were made between the used algorithms which aims to find the best accurate algorithm to predict the fatality of road accidents in Lebanon.

2. DATA AND METHODOLOGY

Road traffic accidents are considered as the most important cause of deaths in Lebanon according to the latest WHO report published in 2017, where Lebanon ranks 86th in the world [14]. As it is noticeable through the graph (Figure.1), the road accidents phenomena have aggravated significantly in Lebanon in the recent years.

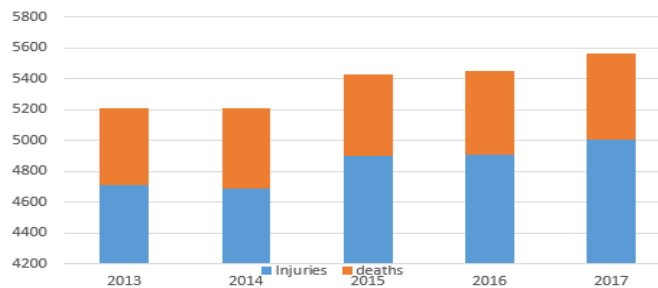


Figure 1 . Aggravation of road accidents in Lebanon
From the year 2013 to 2017

For this reason, it became obvious to propose an approach that can deal with this worst phenomenon. Our process comprises a comparative study of performance of various data mining techniques that predict the fatality of road accidents in Lebanon. The figure below (Figure.2) illustrates a short summarization of the proposed approach.

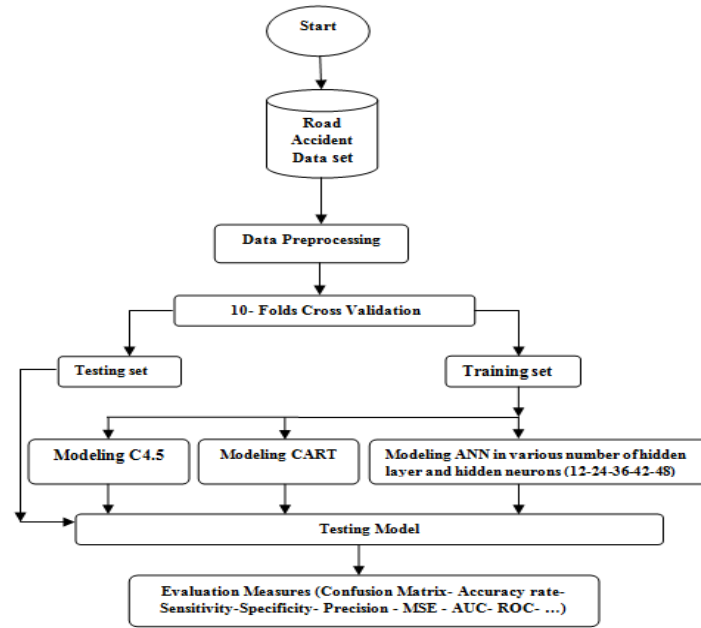


Figure 2. Flow chart of proposed approach

2.1. Data set

Road accidents spread over the Lebanese territory are characterized by diverse natural and human traits. For our approach, we have used the data set collected from the Lebanese ISF for the year 2016-2017. Twelve attributes, figured out throughout the data, are under study to identify the main causes of road accidents in Lebanon. A class attribute is built for the output to select the fatality rate of accidents (“Injured” or “Dead”). The total recordings are 11014 accidents (1100 dead and 9914 injured). The table below describes the nominal values for each attributes.

TABLE 1. List of attributes and their values

Attributes	Values
Month	January, February, March, April, May, June, July, August, September, October, November, December.
Day	Monday, Tuesday, Wednesday, Thursday, Friday, Saturday ,Sunday
Official Holidays	Yes, No
Time	Morning, Afternoon, Evening, Night,
Type	Head-on collisions Multiple vehicle collisions, Rear-end-collisions, Side-impact collisions ,Sideswipe collisions ,Single-car accidents , Vehicle rollover
Cause	Defects, Distracted Driving, Drunk or Drugs Driving, Design Night Driving, Reckless Driving, Running Red Lights, Sliding, Speeding, Tailgating, Teenage Drivers, Wrong-way Driving, Tire Blowouts
Casualty	Driver, Passenger in front, Rear passenger, Pedestrian, bicyclist
Road	Main, International, Internal

Road Type	One way in one direction, One road in two directions, Two ways in two directions
Road Status	Dry, Wet , Ice
Road shape	Straight, Bridge, Cross, Curve, Junction, Slope, Tunnel, Turn
Weather	Clear , Cloudy, Sunny, Rainy, Iced.
Class	Injured, Dead

2.2. Data Pre-processing

Data preprocessing is an important phase in data mining in order to prepare the data set. The data set in its original collected format is not suitable for data mining applications. The R software tool is used in this study for its proven fluency in machine learning applications. We need to convert our original format data to ARFF format file (Attribute-Relation file format). Furthermore, we've made multiplication to the data to obtain balance between dead and injured cases.

2.3. K-folds cross validation

Supervised machine learning techniques formulate the majority of data mining techniques where the process of an algorithm is to learn from the training data set. K- folds cross validation is a procedure used to split the data into k random equal smaller sets as named folds. One of the k sets is retained as testing set and the remainder k-1 as training sets. This process is then repeated for k iterations. Their results can then average to produce a single estimation. For our approach we choose that k=10, where the data are randomly split into 90% for training and 10% for testing for each iteration.

2.4. Data mining applications

Data mining techniques have become a popular tool to reveal a statistical analysis, to evolve a prediction and to explore meaningful patterns and rules out of a large data set. In our study, we develop models of decision trees (C4.5 and CART) and artificial neural networks constituted of various number of hidden neurons.

2.4.1. C4.5 algorithms

C4.5 algorithm is an extension of ID3. C4.5 is a univariate decision tree introduced by Quilan Ross. This algorithm builds a decision tree using the concept of information entropy. For a given probability distribution $P=(p_1, p_2, \dots, p_n)$, Shannon Entropy $E(P)$ and Information Gain are evaluated as follows [15]:

$$\text{Entropy}(P) = E(P) = \sum_{i=1}^n P_i * \log(P_i) \quad (1)$$

$$\text{Gain}(P,T) = \text{Entropy}(P) - \sum_{j=1}^n (P_j * \text{Entropy}(P_j)) \quad (2)$$

Where P_j values are the set of all possible values for attribute T

The information gain measure is used to build the tree by selection of the test features at each node that is ranking the attributes. The attribute with the highest information gain ratio is chosen as the test feature for the current node. [16]

2.4.2. CART algorithm

CART stands for Classification and Regression Trees, proposed by Breiman. CART algorithm is characterized by the fact that it constructs binary trees, namely each node has exactly two outgoing edges [17]. This algorithm is able to generate regression trees where their leaves predict a real number and not a class [18]. Also CART trees are proposed to produce a sequence of nested pruned trees. Whereas in this case CART explores splits that minimize the prediction squared error. For classification tasks, GINI impurity index is used for the splitting procedure.

$$GINIindex = 1 - \sum_j (P_{j/t})^2 \quad (3)$$

Where $(P_{j/t})$ is the relative frequency of class j at node t .

2.4.3. Artificial Neural Network

Artificial Neural Network (ANN) is defined to imitate the computational process of the human brain. Basically a brain is made of cells called neurons. The neurons are interconnected through synapses located on the dendrites. ANN process is a mimic of the biological synapse phenomena. It consists of inputs that are multiplied by weights and then computed by a mathematical function that determines the activation of neurons. Also another function is defined to compute the output of the artificial neuron. Fig.3 represents the architecture of an ANN model. Multilayer Perceptron MLP is the most usable ANN technique, where the information are transmitted in one direction from the input layer to the output passing through hidden layer(s). MLP can be considered actually as a special case of non-linear regression techniques [19].

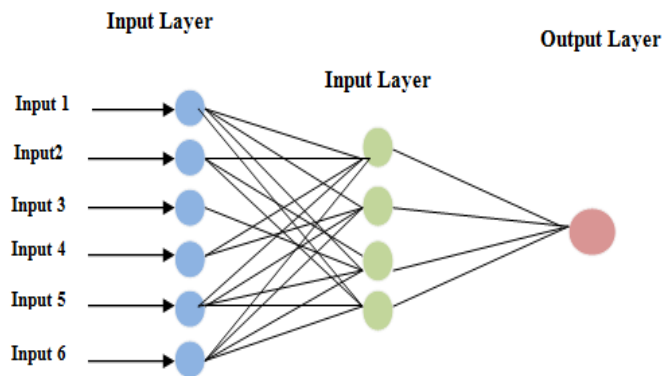


Figure 3. Architecture of an ANN model

3. EXPERIMENTAL RESULTS AND DISCUSSION

In order to build classification models in data mining, R software tool has been used. Lebanese road traffic accident data for 2016 and 2017 are explored in this study. A balancing of data generates 19814 samples with 12 attributes. The classification models are: Decision trees (C4.5 and CART) and artificial neural networks. The splitting of balanced data is executed by applying 10-folds cross validation. This study aims to predict the fatality (Dead or injured) for Lebanese road accidents. A comparative performance evaluation of each model is revealed to select their priority according to various measures: sensitivity, specificity, precision, accuracy, mean squared error (MSE) and area under curve (AUC).

3.1. Decisions trees analysis results

AC4.5 and CART classifier models are experimented in this approach to analyze the data and to evaluate the performance. The experimental results have revealed according to their confusion matrix that 1835 and 1563 accidents have been correctly classified by C4.5 and CART respectively. The figures below (Figure 5 and Figure 6) illustrate the testing confusion matrix of C4.5 and CART.

Where the confusion matrix architecture represents in figure the parameters as follow:

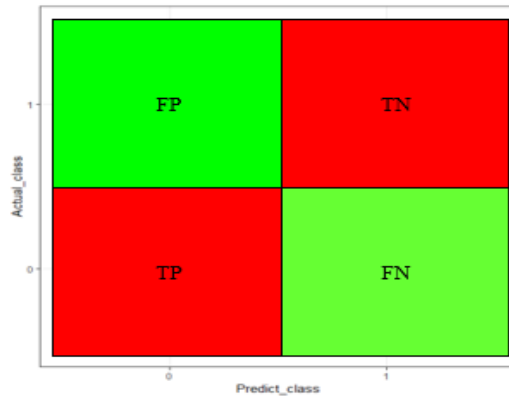


Figure 4. Confusion matrix architecture

- TP = true positive (the network predicts a true fatal road accident)
- FP=false positive (the network predicts a no-fatal road accident).
- TN=true negative (the network predicts a true no-fatal road accident).
- FN=false negative (the network does not predict a true fatal road accident).

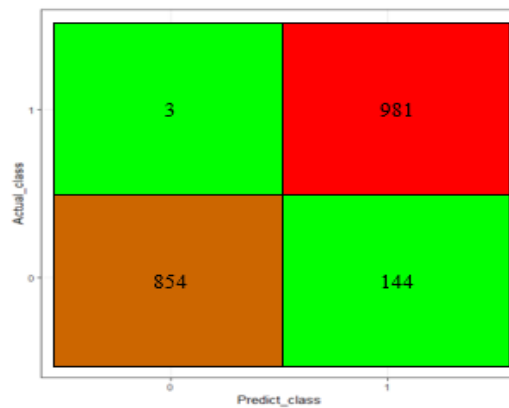


Figure 5.C4.5 Confusion Matrix

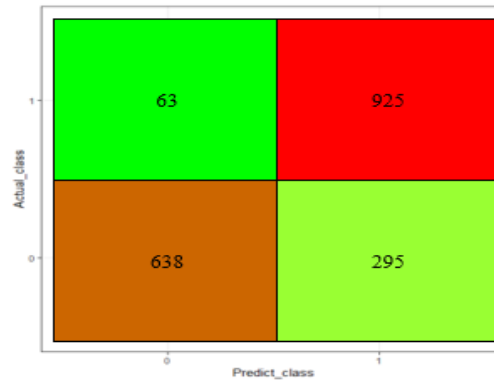


Figure 6. CART Confusion Matrix

Furthermore, CART algorithm has scored 81.3% in terms of accuracy and 85.8% in terms of AUC while C4.5 has revealed an accuracy rate higher than CART algorithm, 92.5%, and AUC of 94.4%. The table below (Table 2) summarizes the values of several measures resulting from the two decision tree models C4.5 and CART. Also The ROC curve of each model is illustrated in the two figures below (Figure.6 and Figure.7).

Where the various measure of performance represented in the Table 2 has as definition:

- Sensitivity = $TP / (TP + FN)$: Probability of correctly labeling members of the target class.
- Specificity = $TN / (TN + FP)$: a statistical measure of how well a binary classification test correctly identifies the negative cases.
- Precision = $TP / (TP + FP)$: Probability that a positive prediction is correct.
- Accuracy = $(TP + TN) / (TP + TN + FP + FN)$.

Table 2. Measure of Sensitivity, Specificity, Precision, Accuracy, MSE and AUC for testing data

Model	Sensitivity %	Specificity %	Precision %	Accuracy %	MSE %	AUC %
CART	68.3	93.6	91	81.3	18	85.8
C4.5	85.5	99.6	99.6	92.5	7.4	94.9

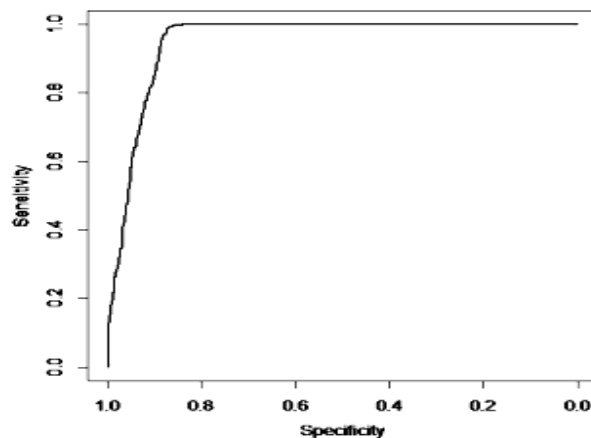


Figure 7. ROC Curve of C4.5 algorithm

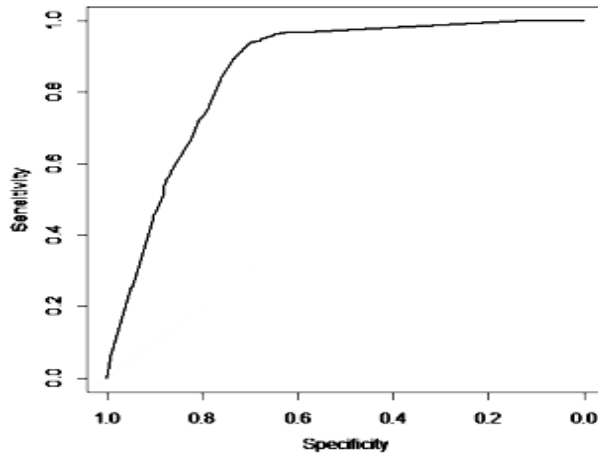


Figure 8. ROC Curve of CART algorithm

According to the experimental results analysis, C4.5 outperformed CART algorithm. Since C4.5 recorded the highest accuracy in predicting such phenomena. In addition to that the area under curve recorded (94.9%). Moreover, the trees generated by these algorithms lead to identify that for rollover and single car accident, speeding and Reckless driving are the most effective factors of fatal accidents in Lebanon.

3.2. Artificial Neural Networks

In the case of applying artificial neural network, MLP models have been experimented with several modifications in the number of hidden layer and number of hidden neurons. Subsequently, MLP models have been trained by standard backpropagation learning function and sigmoid activation function in the output layer synchronized by 1000 as maximum number of iterations to learn. In this approach, we have examined several MLP models by selected 12 inputs according to 12 attributes and 2 outputs representing the output class attribute (Dead and injured). The models are built by changing the hidden neurons in several cases of hidden layer(s) (one and two-hidden layers). The experimental results of performance evaluation of training and testing test for each model are represented in the table below (Table 3).

Table 3. Measure of Sensitivity, Specificity, Precision, Accuracy, MSE and AUC for training and testing data

Sets	Nb. of layers	Nb. of neurons in each layer	Sensitivity %	Specificity %	Precision %	Accuracy %	MSE %	AUC %
Training	1	12	74.2	88.7	86.8	81.4	18.5	87.6
		24	82.7	90.8	90.03	86.7	13.2	91.14
		36	86.9	96.4	96.09	91.6	8.32	82.82
		42	88.2	97.6	97.34	92.9	7.07	93.66
		48	88.9	96.75	96.4	92.8	7.12	93.7
	2	(12,12)	74.2	88.7	86.8	81.4	18.5	90.6
		(24,24)	91.2	95.9	95.7	93.6	6.38	96.08
		(36,36)	94.1	98.3	98.2	96.2	3.73	97.78
		(42,42)	96.02	98.7	98.7	97.4	2.59	98.39
		(48,48)	95.4	98.2	98.1	96.8	3.14	98.16
es	1	12	71.2	88.4	85.7	79.9		86.4

						20.08	
	24	80.9	90.06	88.75	85.5	14.4	89.4
	36	80.89	96.5	95.76	88.7	11.2	90.66
	42	81.2	96.99	96.6	88.9	11.09	90.19
	48	83.03	96.5	96.09	89.6	10.3	90.4
2	(12,12)	74.01	94.2	93.06	83.8	16.1	89.1
	(24,24)	83.6	95.1	94.5	89.4	10.5	92.72
	(36,36)	87.02	97.6	97.4	92.2	7.71	94.5
	(42,42)	90.06	99	98.8	94.6	5.3	95.71
	(48,48)	87.4	98.4	98.3	92.8	7.16	94.56

Accordingly, to the performance evaluation of the application of several MLP models, it is clear that the MLP comprising 2 hidden layers with 42 neurons for each has scored the highest accuracy rates 97.4% and 94.6% for training and testing sets respectively which means MLP (42,42) model has established superiority over other MLPs. While the accuracy of the 12 neurons in 1 hidden layer recorded the worst value (74.2 %)

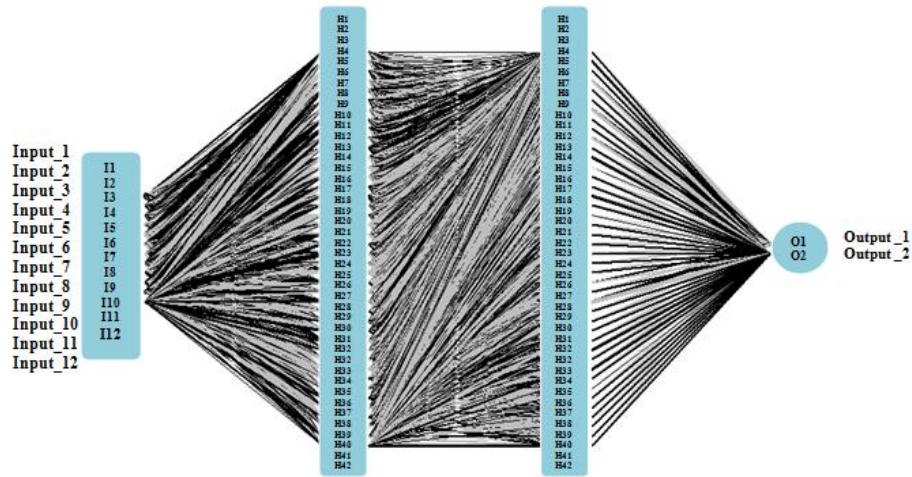


Figure 9. Architecture of MLP (42,42) model

The figure above represents the architecture of MLP (42, 42) model, which consists of 12 inputs and 2 hidden layers that include 42 neurons each and 2 output classes.

The confusion matrix in figure (Fig.) has demonstrated how the MLP (42, 42) model has classified the training and testing sets. This MLP model has revealed that 17831 and 11982 accidents have been correctly classified for training and testing respectively.

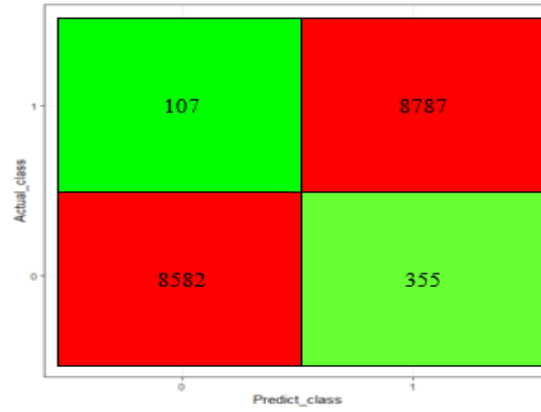


Figure 10. Training confusion matrix

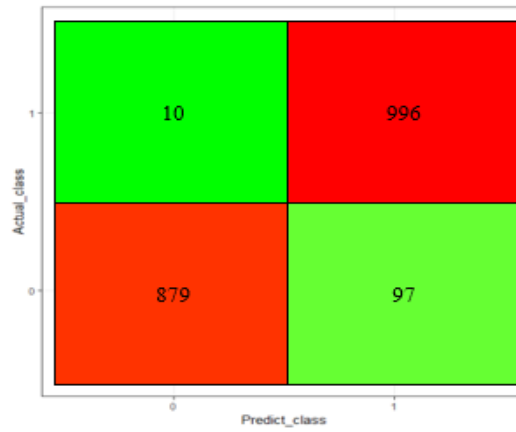


Figure 11. Testing confusion matrix

Moreover, ROC curves of MLP (42, 42) are represented in the figures below (Fig.11 and Fig.12) to clarify the area under curve measure results.

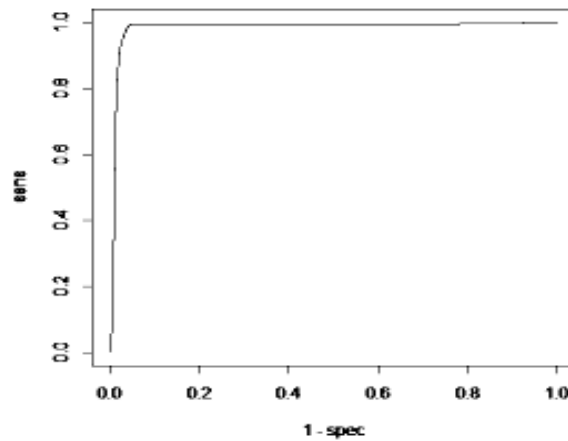


Figure 12. ROC curve of training Data for MLP(42,42) model

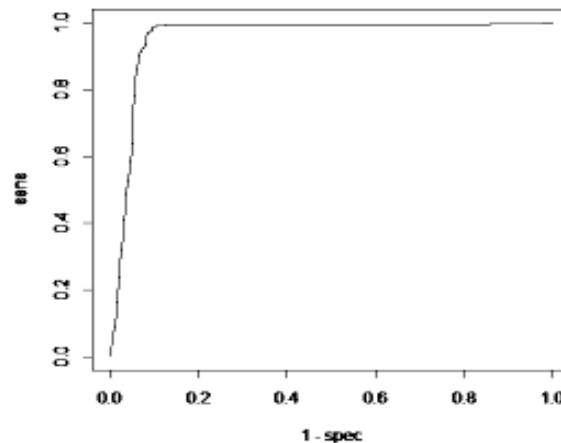


Figure 13. ROC curve of testing Data for MLP(42,42) model

As seen, many measures could be computed for evaluation of the models under study: Accuracy, Precision, Sensitivity, Area Under the Curve, and others. But these shouldn't be used without a clear understanding of the studied field and corresponding identification of chance or base case levels of the study. Using these measures, a system that performs worse in one of the measures, can appear to perform better under any of other commonly used measures.

In the field of road accidents severity prediction, TP and FN tend to be the most important parameters that would affect negatively on the model decision, while FP and TN are less significant. Human beings lives, their properties and the environment are much more valuable than the costs that could be spent on preventive measures in case of false alarms. Thus in our case study, AUC and Sensitivity are found to be the most critical measurements for an adequate evaluation, as both formulas depend on TP and FN.

Hence, MLP (42,42) is considered the best performing model as it retrieves the best sensitivity (90%) and the best AUC (95.7%).

4. CONCLUSION

In this paper, diverse data mining techniques (Decision tree (C4.5 and CART), and various MLP) are employed to determine the main causal factors of road accidents in Lebanon and to find the algorithm that best predicts road accidents severity in the Lebanese territory. The application of these models by using R software data mining tool on Lebanese collected data set has achieved a performance evaluation ranking. MLP (42, 42) model has recorded the best accuracy for testing set (94.6%), highest sensitivity (90%) and highest AUC (95.71%), followed by C4.5 algorithms which has scored 92.5% in terms of accuracy rate, 85.5% in terms of sensitivity and 94.9% in terms of AUC. Otherwise, these classification models identify that, for rollover and single car accidents, speeding and reckless driving are the most important causal factor in fatal road accidents in Lebanon.

In order to validate the obtained results, it is worth working on a bigger database as 2-years dataset is considered imperfect to draw conclusions particularly in the case of handling complicated phenomena such as road accidents. Furthermore, the selected attributes are to be more specific. For instance, reckless driving is considered very wide and each record of accidents shall be reported more specifically that is defining the real cause: driving without due care and attention, being drunk, texting, etc. The selection of attributes shall take into consideration the psychological situation of the driver and his behavioral state. Thus, the mechanism of road

crashes investigating, reporting and recording shall be revised and unified by the officers involved. This will help in keeping track of the actual influencing factors of fatal road accidents and road safety measures.

In the future we suggest to apply different data mining techniques algorithms such as: Artificial Neural Networks deep learning, Support Vector Machine (SVM) and fuzzy logic. These suggestions can lead us to breaks down the nonlinear relationship between accidents attributes and fatal road accident.

ACKNOWLEDGEMENTS

The authors would like to thanks the Lebanese Internal Security forces (ISF) for their kind help in providing us with the Lebanese car accidents data base

REFERENCES

- [1] WHO. Global Status Report on Road Safety 2018: Supporting a Decade of Action. WHO: Geneva
- [2] S. Shanthi and R. Geetha Ramani, "Feature Relevance Analysis and Classification of Road Traffic Accident Data through Data Mining Techniques". Proceedings of the World Congress on Engineering and Computer Science, WCECS 2012, October 24-26, 2012, San Francisco, USA
- [3] L. Li, S. Shrestha, G. Hu, "Analysis of Road Traffic Fatal Accidents Using Data Mining Techniques", IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA), 2017
- [4] S. Krishnaveni, M. Hemalatha, "A Perspective Analysis of Traffic Accident using Data Mining Techniques". International Journal of Computer Applications, 23(7), June 2011
- [5] N. Hamadeh, A. Karouni, B. Daya, P. Chauvet, Comparative Study of Different Data Mining Techniques in Predicting Forest Fire in Lebanon and Mediterranean, Lecture Notes in Networks and Systems, pp: 747-762, 2018
- [6] J. Alzubi, A. Nayyar, A. Kumar, "Machine learning from theory to algorithms: an overview", In Journal of Physics: Conference Series, 23(7), November 2018
- [7] A. Nayyar, V. Puri, "Comprehensive Analysis & Performance Comparison of Clustering Algorithms for Big Data" Review of Computer Engineering Research, 4(2), 54-80, 2017
- [8] T. Kebede Bahiru, D. Kumar Singh, E. Ayalew Tessfaw, "Comparative Study on Data Mining Classification Algorithms for Predicting Road Traffic Accident Severity". Proceedings of the 2nd International Conference on Inventive Communication and Computational Technologies, ICICCT 2018
- [9] I. Al-Turaiki, M. Aloumi, N. Aloumi, and K. Alghamdi, "Modeling traffic accidents in Saudi Arabia using classification techniques. 4th Saudi International Conference on Information Technology (Big Data Analysis) (KACSTIT), pp: 1-5, 2014
- [10] E. Suganya and S. Vijayarani, "ANALYSIS OF ROAD ACCIDENTS IN INDIA USING DATA MINING CLASSIFICATION ALGORITHMS. Proceedings of the International Conference on Inventive Computing and Informatics (ICICI 2017) pp: 1122-1126, 2017
- [11] P. Baluni, Y. P. Raiwani, "Vehicular Accident Analysis Using Neural Network". International Journal of Emerging Technology and Advanced Engineering, Vol:4, Issue 9, September 2014
- [12] V. A. Olutayo, A. A. Eludire, "Traffic Accident Analysis Using Decision Tree and Neural Networks". I.J. Information Technology and Computer Science, Vol: 2, pp: 22-28, 2014

- [13] M. YasinÇodur and A.Tortum, "An artificial neural network model for highway accident prediction: a case study of Erzurum, Turkey". *Promet – Traffic & Transportation*, 27(3), pp: 217-225, 2015
- [14] WHO. WORLD HEALTH RANKINGS : www.worldlifeexpectancy.com
- [15] B. Hssina, A. Merbouha, H. Ezzikouri, M. Erritali, "A comparative study of decision tree ID3 and C4.5". *International Journal of Advanced Computer Science and Applications*, Special Issue on Advances in Vehicular Ad Hoc Networking and Applications, pp: 13-19, July 2014
- [16] S. Sharma, J. Agrawal, S. Sharma, "Classification Through Machine Learning Technique: C4.5 Algorithm based on Various Entropies", *International Journal of Computer Applications* (0975 – 8887) Vol :82 (16), November 2013"
- [17] M.Venkatadri and C.Lokanatha Reddy, "A comparative study on decision tree classification algorithms in data mining". *International journal of computer applications in engineering, technology and sciences*, 2(2), pp: 24-29, 2018
- [18] S. Singh and P. Gupta, "COMPARATIVE STUDY ID3, CART AND C4.5 DECISION TREE ALGORITHM: A SURVEY", *International Journal of Advanced Information Science and Technology (IJAIST)*, 27 (27), July 2014
- [19] G. R.HalagundegowdaandS.Abhishek, "Multilayer Perceptron Method of Artificial Neural Network for Classification of Farmers Based on Adoption of Drought Coping Mechanisms".*Int. J. Pure App. Biosci.* 6 (2), pp: 1408-1414, 2018

Authors

ZeinabFarhat, Eng, has received her degree in computer communication engineering, from the Faculty of Engineering, Lebanese university in 2008. Then she received her master degree in computer Engineer from AUL 2014. She is now a doctoral student at the faculty of computer engineering, Lebanese university EDST.

Ali Karouni, Dr, has received his degree in electrical and electronics engineering, industrial digital control major from the Faculty of Engineering, Lebanese University in 2009. He received his master de-gree in industrial control from the EcoleDoctorale des Sciences et de Technologie -Lebanese University in collaboration with Universite de Technologie de Compiegne-France in 2010. He then was granted a diploma of master in renewable energies from Saint Joseph University-Faculty of Engineering ESIB-Lebanon in collaboration with EcoleDoctorale des Sciences et de Technologie-Lebanese University. He has received his Ph.D. from Universited'Angers-France collaborated with EcoleDoctorale des Sciences

Bassam Daya, Prof, has received his BE degree in electrical and computer engineering in 1992 from the Lebanese University, Lebanon, the MS degree in automatic control and applied computer in 1993 from the Ecole Centrale of Nantes, France and the PhD degree in automatic control and applied computer in 1996 from the University of Angers, France.

Pierre Chavet, Prof, is currently working as a full professor at the Institute of Applied Mathematics of UniversitéCatholique de l'Ouest (Angers, France) and as a director of the research team of the institute. He is also a member of the Laboratoire d'Ingénierie des SystèmesAutomatisés (LISA) research laboratory at Universitéd'Angers. He holds a Ph.D. (1993) in automatics from Universitéd'Angers (France) and an HdR ("accreditation to supervise research" -2001) in computer science from Université Paris 8 (France).