2022 International Conference on Intelligent Systems and Computer Vision (ISCV) | 978-1-6654-9558-5/22/$31.00 ©2022 IEEE | DOI: 10.1109/ISCV54655.2022.9806099

Prediction of Traffic Accidents using Random Forest Model

Imad EL MALLAHI*, Asmae DLIA, Jamal RIFFI, Mohamed Adnane* Mahraz*, and Hamid TAIRI*

Sidi Mohammed ben abdellah University, Faculty of sciences dhar el Mahraz, Department of computer sciences, LISAC laboratory, Fez, Morocco

Abstract. With the increasing trend of the accident rate, the number of casualties in humans has increased considerably over the past decades, which has led to the use of cameras, or fixed speed cameras to carry out their routine activities. In this paper, we focus on severity prediction of traffic accidents, which is a huge step in road accident management in the road. This problem provides important information for emergency logistical transportation in many cities. To evaluate the severity of road accidents in the crowded target, we evaluate the potential impact of the accident to realize effective accident management procedures. In this proposed study, we implement and compare some algorithms in machine learning such as Random Forest, Support Vector Machine, and Artificial Neural Network to classify and predict severity for Traffic accidents, and we presented the confusion matrix to specify the impact of different classes on each other for : Pedestrian, Vehicle or pillion passenger, or Driver or rider to validate this experimentation. In the numerical example we use the TRAFFIC ACCIDENTS_2019_LEEDS data from the Road Safety of department Transport to classify the Severity prediction for Traffic accidents into three classes: Pedestrian, vehicle or pillion passenger, and driver or rider to have a 93% accuracy for Random Forest compared to 82% for SVM and 87 for ANN, and at the level of precision recall we also have 93.82% for Random Forest compared to 82.22% for SVM and 87.88% for ANN.

Keywords: Machine learning, Traffic accident, severity prediction, convolutional neural network

1 INTRODUCTION

Recently, with the increasing trend of the accident rate, the number of casualties in humans has increased considerably over the past decades, which has led to the use of cameras, or fixed speed cameras to carry out their routine activities. the issue of the severity prediction for Traffic accidents is the big interests in the world [1-5]. The proposed work aims to provide a prediction tool for the problem prediction the severity of the important information for emergency logistical [6-7]. The most challenge is the lack of real-time data from the Road Safety for the traffic accidents.

The proposed work performed statistical significance testing on the impact of applied a multi-class neural network and multiclass random forest on a Traffic accidents Data Set [8-12]. Some algorithm of Machine learning can help in complicated decision supports system solutions [13-22], also some authors discuss the issue of traffic light control is a challenging problem in modern societies [23-27]. The paper presented an efficient solution to use data to severity prediction by detecting the severity prediction issue for Traffic accidents. In this paper two machine learning techniques were proposed for the detection of severity prediction for Traffic accidents. The multi-class neural network proved a better accuracy with 93.64% to predict severity prediction more than the multi-class random forest which achieved 87.71%, in the author hand the Random Forest Algorithm combines the output of multiple (randomly created) Decision Trees to generate the final output . Conclusion: Applying machine learning algorithms on severity prediction data can help severity prediction providers and individuals to pay attention to the Traffic accidents risks and Traffic accidents status changes to improve the quality of life. The proposed system was applied to a Traffic accidents Data Set. The experimental results of the proposed work proved that using the multi-class neural network method can increase the possibility of diagnostic accuracy.

2 Related Work

During the last ten decades, the issue of road traffic safety has a big interest for economic and social development in the world. The various solution in intelligent systems were employed for traffic accidents classification. Fig. 1 represents the road traffic accident car crash, Earlier methods used manually defined features, mostly based on the combination of images of accidents in the road, and statistical information [28].

Fig. 1. road traffic accident car crash

During this year 2022, in Morocco the Speed cameras fixed in the roads to control and transmit images and videos in real time in the event of traffic violations or accidents, which can help in the decision for road accidents, Fig. 2. Represent the new generation radars commissioned in 2022 in Morocco .

Fig. 2. New generation radars

In 2022, 522 new radars will be made available to National Security and Royal Gendarmerie Road control officers. The new generation mobile speed cameras will allow speed control, overtaking and compliance with red lights.

To fight against speeding which is one of the main causes of death on the roads, the authorities want to further strengthen road control, Morocco will increase the number of traffic cameras from 120 to 672 in 2022. This information has been reported by the National Road Safety Agency (NARSA). These are intelligent cameras and radars capable of recording several types of violations. Reaching the number of 672 speed cameras in 2022 is against an initial target of 1,200. The investment cost is estimated at 278.5 million dirhams. With regard to fixed speed cameras, which are a total of 552 units, they will be used to monitor speed and respect for red lights inside and outside urban areas. Their installation will be done right after the completion of the operation of locating the installation locations.

These radars include 72 average speed control devices on motorways, 276 speed control devices outside urban perimeters, 204 control devices for respecting speed traffic lights in urban perimeters. As for the speed control devices, they will be installed at the level of the motorway network and will make it possible to control the average speed of vehicles at the level of the motorway sections over several kilometres.

The particularity of these new radars is that in addition to detecting violations relating to speeding and traffic lights, they can detect more than one vehicle at a time, up to 24 vehicles. They have several other characteristics, such as the detection of infringements relating to the continuous line, the distinction between light vehicles and heavy goods vehicles, the possibility of operation in both directions (distance-approach) and the use of the laser scanning technique.

With the arrival of the new radars, the number of regions with automated traffic monitoring will increase from seven to twelve. Distributed in descending order among the 12 regions, 108 will be placed in Casablanca-Settat, 92 in Marrakech-Safi, 69 in Rabat-Salé-Kénitra, 62 in Fès-Meknes, 51 in Tangier-Tétouan-Al Hoceïma, 45 in the Oriental, 43 in Béni Mellal-Khénifra, 38 in Souss-Massa, 13 in Drâa-Tafilalet, 11 in Guelmim-Oued, 10 in Laâyoune-Sakia El Hamra and finally 10 in Dakhla-Oued Ed-Dahab.

Manually defined and feature descriptors are fed to the traditional machine learning models, SVM [29]. The comprehensive survey on the studies of computer vision approaches for traffic accidents recognition can be found in [30].

With the advances of hardware as speed cameras fixed, especially with the incorporated use of GPUs, deep neural networks (DNNs) have achieved new standards in many research frontiers. The main advantage of the DNNs is that they do not require manual feature selection, and the features are learnt within a DNN framework. However, DNNs require a large amount of training data, which is not always available. For small or moderate size datasets, transfer learning can help to overcome dataset size limitation [31]. Delen D, Sharda. [32] identified the significant predictors of injury severity in traffic accidents using a series of artificial neural networks. Alikhan, and Lee [33-34] using the clustering-classification heuristic method for improvement accuracy in classification of severity of road accidents .

The aim of this paper is threefold. First, we introduce a TRAFFIC ACCIDENTS_2019_LEEDS dataset. Second, we analysed the quality of this dataset for the traffic accident classification task. Third, we extend further the study using ANN, SVM, and random forest model to pre-training for traffic accident classification task by exploring a larger number of machine learning models.

3 PROPOSED METHOD

3.1 Dataset employed

In this proposed study, we use then TRAFFIC ACCIDENTS_2019_LEEDS data from the Road Safety of department Transport. The label of classification has represented to each of the data. In this database there are 1152 were classified as Pedestrian, 405 were classified as Driver or rider and the remaining 350 were classified as Vehicle or Pedestrian passenger.

Table 1. Complete Dataset details

Type	Number of features
Pedestrian	**1152**
Driver or rider	**405**
Vehicle or Pedestrian passenger	**350**
Total	**1907**

Table 1 presents the dataset details, and number of features for Pedestrian, Vehicle or pillion passenger, or Driver or rider.

3.2 Balancing the database

As shown in the following Fig. 3., the database is unbalanced, because the number of each class is quite different Pedestrian, Vehicle or pillion passenger, or Driver or rider.

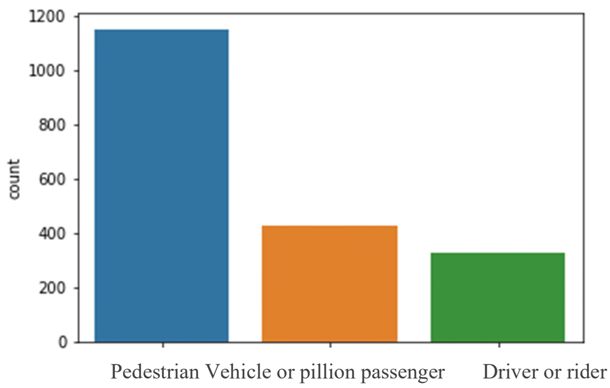


Fig. 3. Distribution of casualty class for the severity prediction for traffic accidents.

To balance the database, there are two possibilities, Up sampling: resample the values to make their count equal to the class label with the higher count • Down sampling: pick n samples from each class label where n = number of samples in class with least count (here, 128) In this study, we chose to expand the database. We obtained 1152 records for each class, for a total of 3456 records after augmentation.

Table 2. Dataset details after **augmentation**

Type	Number of features
Pedestrian	1152
Vehicle or pillion passenger	1152
Driver or rider	1152
Total	3456

Then we divided the database into two parts, a training part (Training Dataset) and another part for testing (Test Dataset). We used 80% of the database for training and 20% for testing. i.e., 2764 number of features for Pedestrian, Vehicle or pillion passenger, or Driver or rider for training set and 692 features for test set in Table 2.

3.3 Artificial Neural Network proposed

In this experimentation, we took an ANN that consists of an input layer (with 128 neurons), a hidden layer (with 64 neurons) and an output layer (with 3 neurons) with Fig.4 represents the pipeline of the proposed work.

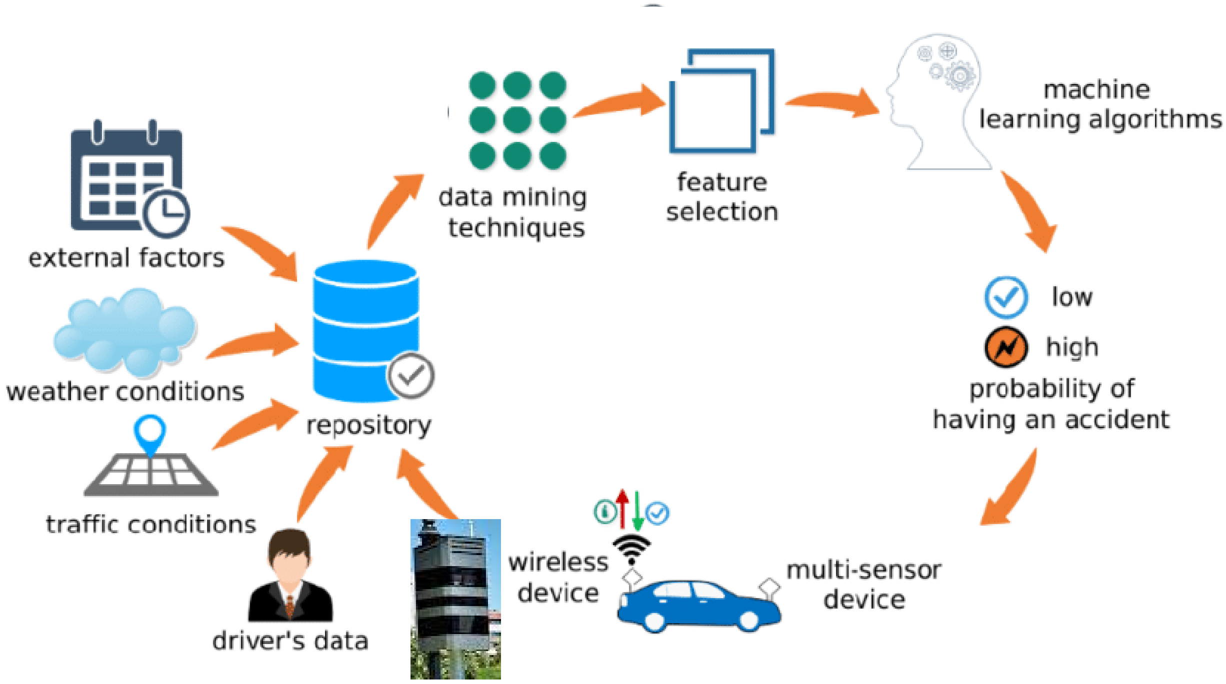


Fig. 4. Methodological steps

4 EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Evaluation metrics

Accuracy: Accuracy is one criterion for evaluating classification models. Informally, accuracy refers to our model's percentage of true predictions. The formal definition of accuracy is as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) are the acronyms for True Positives, True Negatives, False Positives, and False Negatives, respectively

Precision The percentage of successfully detected positives in relation to all expected positives. Mathematically:

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Where TP denotes True Positive (number of correct positive predictions) and FP denotes False Positive (quantity of misclassified positive predictions) Recall The total number of positive predictions that were correct across all positive samples. Mathematically:

$$Recal = \frac{TP}{TP + FN} \quad (3)$$

Where TP denotes True Positive (number of correct positive predictions) and FN denotes False Negative (number of incorrect negative predictions)

F1 score Precision and Recall in a symbiotic relationship. For unbalanced data, the F1 score is a superior performance statistic than the accuracy metric.

$$F_1 = 2 \times \frac{Precsion \times Recall}{Precsion + Recall} \quad (4)$$

4.2 Experimental setting

We compared the performance of our model to the performance of the ANN, SVM, and RF approaches.

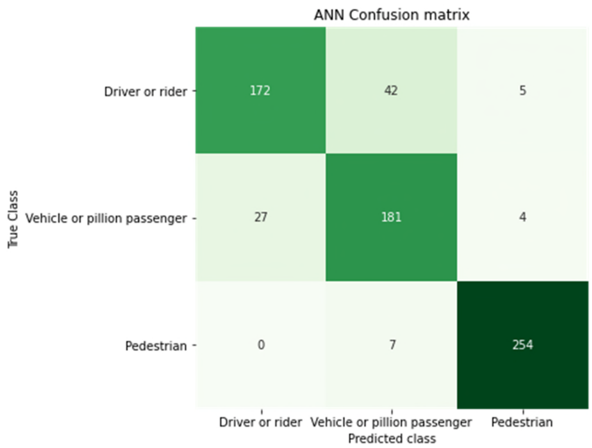
While preserving the class ratio between Pedestrian, Vehicle or pillion passenger and Driver or rider, where the dataset were randomly split into training and test data. Each of the models we tested was trained using trained data,while the models' performance was evaluated using test data. To ensure that the model was consistent, we ran 10-fold cross-validation on each of the models.

To compare results with our system, we used the ANN,SVM, and RF classifiers on TRAFFIC ACCIDENTS_2019_LEEDS dataset . The algorithms were created utilizing the Python scikit-learn toolkit and the hyperparameter settings provided.

Fig. 5. Represents the confusion matrix for Pedestrian, Vehicle or pillion passenger, or Driver or rider using ANN model, the performance of the ANN model for the test dataset is evaluated after the completion of the training phase and was compared using four performance measures such as precision, sensitivity or recall, specificity, precision (PPV), area under the curve (AUC), F1 score. Fig. 5 also presents the confusion matrix for Pedestrian, Vehicle or pillion passenger, or Driver or rider classification using Random Forest model and SVM model. To measure the performance of the model that Fig.5 represents the Train and validation accuracy curve.

$$CM = \begin{bmatrix} TP & FP \\ FN & TN \end{bmatrix} \quad (5)$$

Fig. 5. Confusion matrix for Pedestrian, driver or rider, and vehicle or Pedestrian passenger classification using ANN model



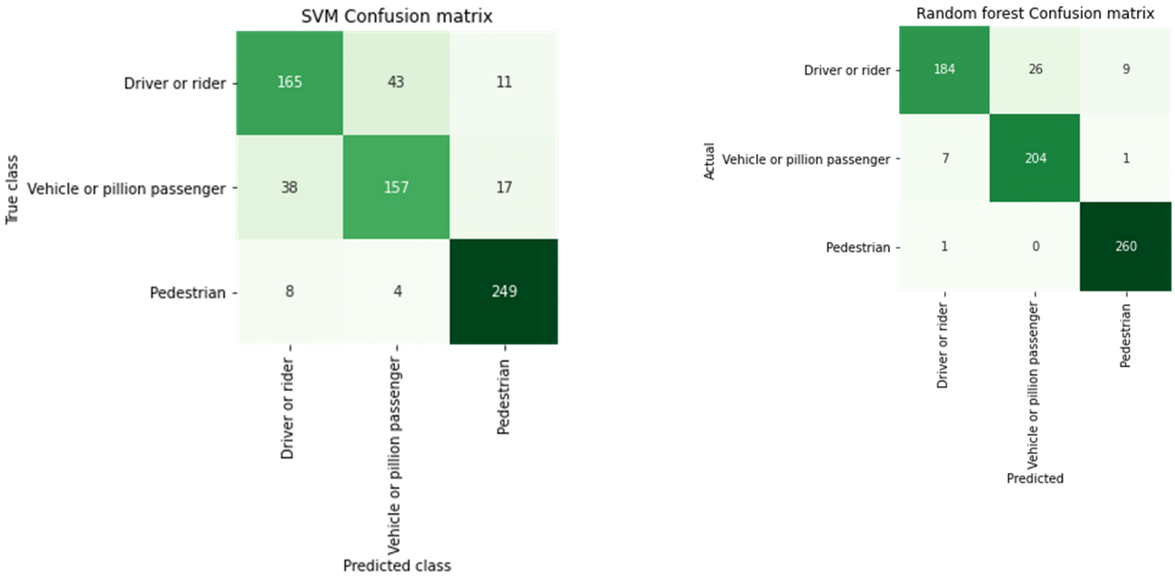


Fig. 5. Confusion matrix for Pedestrian, Vehicle or pillion passenger, or Driver or rider classification using Random Forest model (a) SVM model and (b) Random Forest model.

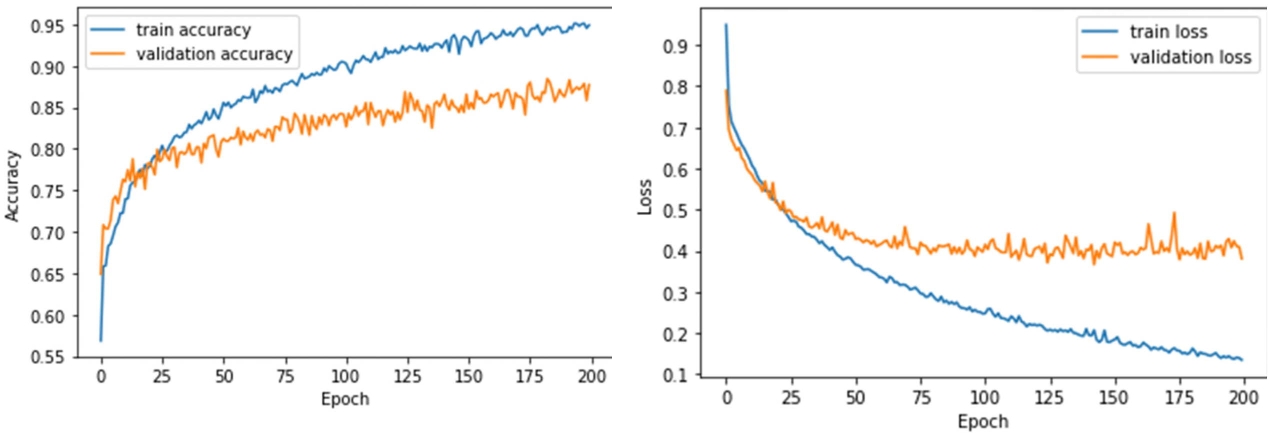


Fig. 6. Train and validation accuracy curve (a) Train and validation accuracy curve, (b) Train and validation loss curve

Table 3. Values obtained for the different metrics

	Random Forest	SVM	ANN
Accuracy	0.9364161849710982	0.8251445086705202	0.8771676300578035
precision	0.9382125952919493	0.8222978546756788	0.8788867858874835
Recall	0.9364161849710982	0.8251445086705202	0.8771676300578035
F1 score	0.9355102879655588	0.8232424765175214	0.8770074547672448

As shown in Table 3, the best performing model for detecting the fatal state is the Random Forest model which gave better accuracy values (93.64% for training accuracy and 93.82% for test accuracy)

5 DISCUSSION

To predict the severity of traffic accidents, we used ANN, SVM and RF classifiers. The results of the models in terms of accuracy, precision, recall and F1 score are calculated from the confusion matrices.

In terms of accuracy, precision, recall and F1 score, Table 3 shows the results of the models on the dataset. Fig. 6 represents the train and validation accuracy curve .

For this dataset, the Random Forest classifier outperforms the other models in terms of precision, accuracy, and F1 score. Although the ANN classifier has the best recall, it performs poorly on the other performance criteria for this dataset. Compared to RF, ANN and SVM classifiers perform admirably. Compared to RF, ANN and SVM are less accurate. However, compared to the other approaches, SVM fails to achieve a satisfactory F1 score and recall score, even though the precision score is correct compared to RF and ANN classifiers.

6 CONCLUSION

In this work, we have focused on severity prediction for traffic accidents, which is a huge step in road accident management. This issue provides important information for emergency logistical transportation. To evaluate the severity of road accidents, we have evaluated the potential impact of the accident, and realized effective accident management procedures. In this proposed study, we have implement some algorithms to classify the severity of Traffic accidents, and presented the confusion matrix to specify the : Pedestrian, Vehicle or pillion passenger, or Driver or rider using Random Forest, Support Vector Machine, and Artificial Neural Network. To validate this experimentation, the TRAFFIC ACCIDENTS_2019_LEEDS dataset are used. To classify the Severity prediction for Traffic accidents into three classes: Pedestrian, Vehicle or pillion passenger, or Driver or rider. In future work, it is possible to use more features and find best features for classifications for real data in our city. Again, we can extract these selected features from program file, also we can implement the cost for the prediction the gravity of Traffic Accidents.

.

References

1. Road Traffic Injuries. Accessed: Jul. 18, 2018. [Online]. Available: http://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries
2. F. Zong, H. Xu, and H. Zhang, ``Prediction for traffic accident severity: Comparing the Bayesian network and regression models," Math. Problems Eng., vol. 2013, nos. 23, 2013, Art. no. 475194.
3. R. Collobert and J. Weston, ``A unied architecture for natural language processing: Deep neural networks with multitask learning," in Proc. ACM 25th Int. Conf. Mach. Learn., 2008, pp. 160-167.
4. J. Dean et al., ``Large scale distributed deep networks," in Proc. Int. Conf. Neural Inf. Process. Syst., 2012, pp. 12231231.
5. L. Deng, G. Hinton, and B. Kingsbury, ``New types of deep neural network learning for speech recognition and related applications: An overview," in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), May 2013, pp. 85998603.
6. K. He, X. Zhang, S. Ren, and J. Sun, ``Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. pattern Recognit., Jun. 2016, pp. 770778.
7. N. Majumder et al., ``Deep learning-based document modeling for personality detection from text," IEEE Intell. Syst., vol. 32, no. 2, pp. 7479, Mar. 2017.
8. A. Severyn and A. Moschitti, ``Learning to rank short text pairs with convolutional deep neural networks," in Proc. 38th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr., 2015, pp. 373382.
9. L. Wang, Y. Li, and S. Lazebnik, ``Learning deep structure-preserving image-text embeddings," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2016, pp. 50055013.
10. O. Abdel-Hamid, A.-R. Mohamed, H. Jiang, L. Deng, G. Penn, and D. Yu, ``Convolutional neural networks for speech recognition," IEEE/ACM Trans. Audio, Speech, Language Process., vol. 22, no. 10, pp. 15331545, Oct. 2015.
11. O. Abdel-Hamid, A.-R. Mohamed, H. Jiang, and G. Penn, ``Applying Convolutional Neural Networks concepts to hybrid NN-HMM model for speech recognition," in Proc. IEEE Int. Conf. Acoust., Speech Signal Process., Mar. 2012, pp. 42774280.
12. A. Krizhevsky, I. Sutskever, and G. E. Hinton, ``ImageNet classication with deep convolutional neural networks," in Proc. Int. Conf. Neural Inf. Process. Syst., 2012, pp. 10971105.
13. Q. Mao, M. Dong, Z. Huang, and Y. Zhan, ``Learning salient features for speech emotion recognition using convolutional neural networks," IEEE Trans. Multimedia, vol. 16, no. 8, pp. 22032213, Dec. 2014.
14. M. Oquab, L. Bottou, I. Laptev, and J. Sivic, ``Learning and transferring mid-level image representations using convolutional neural networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2014, pp. 17171724.
15. P. Swietojanski, A. Ghoshal, and S. Renals, ``Convolutional neural networks for distant speech recognition," IEEE Signal Process. Lett., vol. 21, no. 9, pp. 11201124, Sep. 2014.
16. S. Ji, W. Xu, M. Yang, and K. Yu, ``3D convolutional neural networks for human action recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 1, pp. 221231, Jan. 2013.
17. A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. FeiFei, ``Large-scale video classication with convolutional neural networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2014, pp. 17251732.
18. S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, ``Face recognition:A convolutional neural-network approach," IEEE Trans. Neural Netw.,vol. 8, no. 1, pp. 98113, Jan. 1997.
19. M. G. Karlaftis and E. I. Vlahogianni, ``Statistical methods versus neural networks in transportation research: Differences, similarities and some insights," Transp. Res. C, vol. 19, no. 3, pp. 387399, Jun. 2011.
20. A. S. Al-Ghamdi, ``Using logistic regression to estimate the inuence of accident factors on accident severity," Accident Anal. Prevention, vol. 34, no. 6, pp. 729741, 2002.
21. M. Bédard, G. H. Guyatt, M. J. Stones, and J. P. Hirdes, ``The independent contribution of driver, crash, and vehicle characteristics to driver fatalities," Accid Anal. Prevention, vol. 34, no. 6, pp. 717727, 2002.
22. K. M. Kockelman and Y. J. Kweon, ``Driver injury severity: An application of ordered probit models," Accident Anal. Prevention, vol. 34, no. 3,pp. 313321, 2002.

23. Mohamed AbdElAziz, Khamis Ahmed, El-Mahdy Ahmed El-Mahdy, Kholoud Osama Shata Kholoud Osama Shata, Walid Gomaa Walid Gomaa, "System and Method for During Crash Accident Detection and Notification", Patent Number: 2020/771, Filing Date: 9 June 2020, Filing Place: Egypt.
24. Mohamed AbdElAziz Khamis, Ahmed El-Mahdy, Kholoud Osama Shata. "An in-Vehicle System and Method for During Accident Detection without being Fixed to Vehicle", Patent Number: 2020/769, Filing Date: 9 June 2020, Filing Place: Egypt.
25. Mohamed A. Khamis, Walid Gomaa, Adaptive multi-objective reinforcement learning with hybrid exploration for traffic signal control based on cooperative multi-agent framework, Engineering Applications of Artificial Intelligence, Volume 29March, 2014 pp 134–151https://doi.org/10.1016/j.engappai.2014.01.007.
26. Mohamed AbdElAziz Khamis, Walid Gomaa, Enhanced Multiagent Multi-Objective Reinforcement Learning for Urban Traffic Light Control, Proc. of the 11th IEEE International Conference on Machine Learning and Applications (ICMLA 2012), Boca Raton, Florida, USA, 12-15 Dec. 2012, pp. 586-591.
27. Mohamed A. Khamis , Walid Gomaa, Hisham El-Shishiny, Multi-Objective traffic light control system based on Bayesian probability interpretation, Proc. of 15th IEEE Intelligent Transportation Systems Conference (ITSC 2012), Anchorage, Alaska, USA, 16-19 Sept. 2012, pp. 995–1000.
28. Zhang XG. Introduction to statistical learning theory and support vector machines. Acta Automat Sinica 2000; 26: 32–41.
29. Yuan F and Cheu RL. Incident detection using support vector machines. Transp Res Part C Emerg Technol 2003; 11: 309–328
30. Sharma B, Katiyar VK and Kumar K. Traffic accident prediction model using support vector machines with Gaussian Kernel. In: Pant M, Deep K, Bansal JC, et al. (eds) Proceedings of fifth international conference on soft computing for problem solving, vol. 437. Singapore: Springer 2016, pp.1–10.
31. Flores MJ, Armingol JM and de la Escalera A. Real-time warning system for driver drowsiness detection using visual information. J Intell Robot Syst 2010; 59: 103–125.
32. Delen D, Sharda R and Bessonov M. Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks. Acc Anal Prevent 2006; 38: 434–444.
33. Alikhani M, Nedaie A and Ahmadvand A. Presentation of clustering-classification heuristic method for improvement accuracy in classification of severity of road accidents in Iran. Safe Sci 2013; 60: 142–150. 10.
34. Lee S-L. Predicting traffic accident severity using classification techniques. Adv Sci Lett 2015; 21: 3128–3131.