VOLUME 16, N° 4 2022
Journal of Automation, Mobile Robotics and Intelligent Systems

A Distributed Big Data Analytics Model for Traffic Accidents Classification and Recognition based on SparkMlLib Cores

Submitted: 21st June 2022 ; accepted 2nd August 2022

Imad El Mallahi, Jamal Riffi, Hamid Tairi, Abderrahamane Ez-Zahout, Mohamed Adnane Mahraz

DOI: 10.14313/JAMRIS/4-2022/34

Abstract:
This paper focuses on the issue of big data analytics for traffic accident prediction based on SparkMllib cores; however, Spark's Machine Learning Pipelines provide a helpful and suitable API that helps to create and tune classification and prediction models to decision-making concerning traffic accidents. Data scientists have recently focused on classification and prediction techniques for traffic accidents; data analytics techniques for feature extraction have also continued to evolve. Analysis of a huge volume of received data requires considerable processing time. Practically, the implementation of such processes in real-time systems requires a high computation speed. Processing speed plays an important role in traffic accident recognition in real-time systems. It requires the use of modern technologies and fast algorithms that increase the acceleration in extracting the feature parameters from traffic accidents. Problems with overclocking during the digital processing of traffic accidents have yet to be completely resolved. Our proposed model is based on advanced processing by the Spark MlLib core. We call on the real-time data streaming API on spark to continuously gather real-time data from multiple external data sources in the form of data streams. Secondly, the data streams are treated as unbound tables. After this, we call the random forest algorithm continuously to extract the feature parameters from a traffic accident. The use of this proposed method makes it possible to increase the speed factor on processors. Experiment results showed that the proposed method successfully extracts the accident features and achieves a seamless classification performance compared to other conventional traffic accident recognition algorithms. Finally, we share all detected accidents with details onto online applications with other users.

Keywords: *Big data, machine learning, traffic accident, severity prediction, convolutional neural network*

1. Introduction

Creating communication approaches between vehicle, radar, and computer technology is one of the most critical tasks in modern artificial intelligence. One of the easiest ways for a user to enter information is through traffic accident. Therefore, data analysis processing technology and its processing tools have become a necessary part of the information society. In addition, traffic accident recognition is an essential research aspect of data processing and a vital vehicle-radar–computer interaction technique. Traffic accident data contain semantic and personal characteristics, and environmental information.

Recently, the issue of severity prediction for traffic accidents has become a major concern [1-5]. The study presented here aims to provide a prediction tool for the problem of severity, which is important information for emergency logistics [6-7]. The biggest challenge is the lack of real-time data from road safety departments for the traffic accidents.

The proposed work performed statistical significance testing on the impact of applying a multi-class neural network and multiclass random forest on a traffic accidents data set [8-12]. Some algorithms of machine learning can help in complicated decisions supporting system solutions [13-22]; also, some authors discuss the issue of traffic light control as a challenging problem in modern societies [23-27,36].

This paper presents an efficient solution: to use data in 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 to have better accuracy, with 93.64% accurate severity prediction; this is more than the multi-class random forest, which achieved 87.71% accuracy. 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 accident risks and traffic accidents status changes to improve the quality of life. The proposed system was applied to a traffic accident 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. We use the Apache Spark that supports DL and other big data analytic platforms that support machine learning, such as Hadoop, AzureML, and BigML (Fig.1). DL is a branch of machine learning that can solve classification, prediction, and clustering problems in Internet of Vehicles environments.

Big Data Storage: In order to store the incoming data in real time, we use a special cluster such as HDFS (Hadoop Distributed File System) or any other NoSQL database. We specify for our treatments the data locality procedure by calling from the external requesting system any lamBig Data Analytic function to realize the next step (see Fig. 1). For processing and advanced processing, we can directly apply Map Reduce by translating all processing of classification or recognition to Mappers Tasks and Reducers Tasks. And we have to call directly on the pre-implemented advanced function and APIs from the Apache Spark core SparkMlLib.

Visualization: Here, our eyes' tendency is to get attracted more towards visuals rather than written content. We have several ways to practice this step (see Fig. 2). We first plot the trajectories of vehicles or we can plot histograms.

2. Related Work

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

During this year (2022), in Morocco speed cameras were 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 decisionmaking about road accidents. Manually defined and feature descriptors are fed to the traditional machine learning models, SVM [29]. Comprehensive survey of the studies of computer vision approaches for traffic accidents recognition can be found in Sharma et al. [30].

With the advances in hardware for speed cameras, 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 learned 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 limitations [31]. Delen and Sharda [32] identified the significant predictors of injury severity in traffic accidents using a series of artificial neural networks. Alikhan and Lee [33-41] used the clustering-classification heuristic method for improvement accuracy in classification of severity of road accidents.

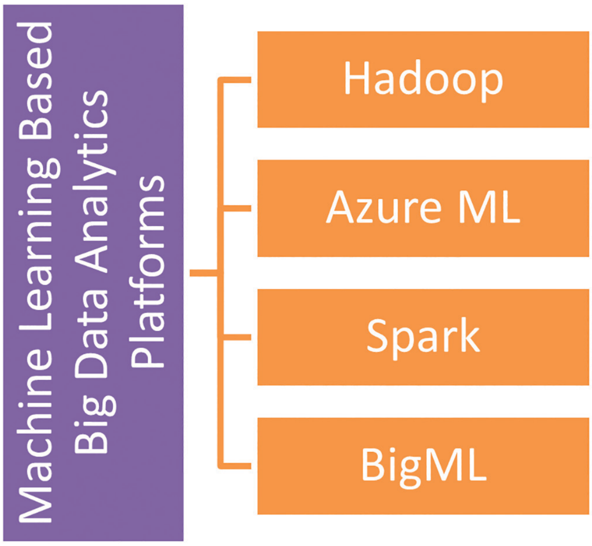

Fig. 1. Big data analytic platforms for machine learning

3. Background

The processing of the large-scale data generated from the Internet of Vehicles environment from various sources, such as cameras and sensors, is required. DL can be used for the processing of the Internet of Vehicles big data. Big Data Analytic platforms that support DL are required for the analysis of Internet of Vehicles.

In this section, we present Apache Spark, which supports ML and other Big Data Analytic platforms that support machine learning, such as Hadoop, AzureML, and BigML. DL is a branch of machine learning that can solve classification, prediction, and clustering problems in Internet of Vehicles environments.

3.1. Apache Spark

Spark is a big data processing framework based on streaming, machine learning, and graph processing

Fig. 2. Big Data processing

[36]. It is an open-source framework and was developed to overcome some of the limitations of Hadoop MapReduce. Spark uses memory based on processing large amounts of data, and it is faster in terms of data processing than the MapReduce framework. As a result, the data are stored in memory using resilient distributed datasets. Moreover, Spark supports real-time analysis. Chiroma et al. [36] presented Spark's open-source distributed machine learning library, MLlib. Several learning settings exist in MLlib to improve the functionality efficiently, such as optimization, linear algebra primitives, and underlying statistical methods. Moreover, MLlib provides a high-level API and several languages that leverage Spark's rich ecosystem to simplify the development of end-to-end machine learning pipelines. Chiroma et al. [36] discussed the DL over Apache Spark for mobile BDA. The authors showed how Spark can perform distributed DL on Map-Reduce. Each partition of the deep model is learned by the Spark worker for the entire mobile big data. Then the parameters use the master deep model of all partial models through averaging.

3.2. Hadoop

Hadoop has emerged as an important framework for "distributed processing of large datasets across clusters of machines" [36]. Many Hadoop-related projects have been developed over the years to support the framework, such as Hive, Pig, Tez, Zookeeper, and Mahout. Mahout is one of the distributed linear algebra frameworks for scalable machine learning.

3.3. AzureML

AzureML is a collaborative machine learning platform based on predictive analytics in big data, which allows easy development of predictive models and APIs. Numerous unique features, such as easy operationalization, versioning collaboration, and integration of user code, are provided by AzureML. Chiroma et al. [36] offered a technique for cloud-based AzureML named Generalized Flow, which allows binary classification and multiclass datasets and processes them to maximize the overall classification accuracy. The performance of the technique is tested on datasets based on the optimized classification model. The authors used three public datasets and a local dataset to evaluate the proposed flow using the classification. The result of the public datasets has shown an accuracy of 97.5%. Furthermore, the concept has become indispensable in big data technologies. For example, AzureML supports neural network for regression, two-class classification, and multiclass classification.

3.4. BigML

BigML provides highly scalable ML and predictive analysis services in the cloud The goal of BigML is to assist in developing a set of services, given that it is easy to use and seamless to integrate. BigML has been used in many studies for predictive analytics and DL because of its robustness and simplicity in providing a user-friendly interface. For example, a study on the distinguishing features of human footprint images offers deep analysis using BigML. The idea is to exploit the concept of the human footprint for personal identification using many fuzzy rules for predictive analysis. The verification of 440 footprint images is conducted for data quality. GPUs have been applied to speed up the performance. Moreover, Chiroma et al. [36] presented a predictive analysis on the most popular place for dengue in Malaysia to obtain an early warning and awareness to people using the BigML platform. The study is based on the decision tree algorithm model, which builds on BigML to support classification. Moreover, Chiroma et al. [36] analyzed the game features and acquisition, retention, and monetization strategies as primary drivers of mobile game application success.

4. Proposed Method

4.1 Dataset Employed

In this paper, we proposed solution-based big data and machine learning models for the development of an intelligent system for traffic accident prediction based on different data sources. Figure 3 represents the networks of wireless access technology involving vehicles and the Internet, as well as the heterogeneous network commonly referred to as the Internet of Vehicles. The figure shows the representation of the Internet of Vehicles in a large-scale distributed environment in terms of wireless communication of various devices. The model of the Internet of Vehicles is integrated into the cloud, equipped with a high-performance computing server with multiple GPUs, large-scale ML models, and Apache Spark. In the first pillar, this processing is done by capturing on real time all incoming datasets, which are stored in the Hadoop Distributed File System cluster. After that we have called on SparkMlLib core to use all the pre-implemented LamBig Data Analytic functions. In the second pillar, an analytical study is performed to group the important features. In the third pillar, we performed a selection of features itself, and new data sets are generated. Fourth, machine learning algorithms are used to define accident rates. Finally, the crash rates are sent to the vehicles. This paper focuses on the second stage of the scheme. The data sources represented for this system are police consulter, traffic conditions, automatic radar, vehicle data, fixed or moving cameras, driver data, weather, or other external factors. Each source for the dataset can be integrated into the proposed system.

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

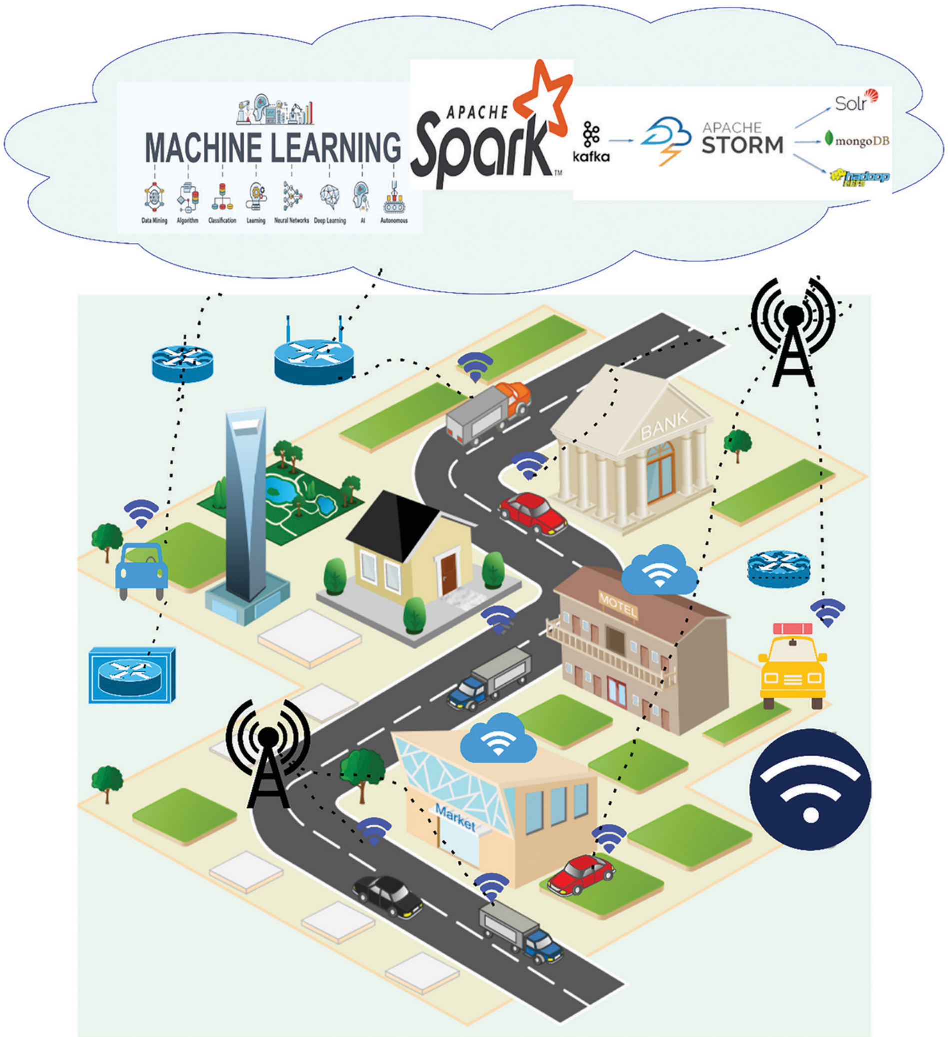

Fig. 3. Model of the Internet of Vehicles integrated into the cloud equipped with a high-performance computing server with multiple GPUs, large-scale ML models, and Apache

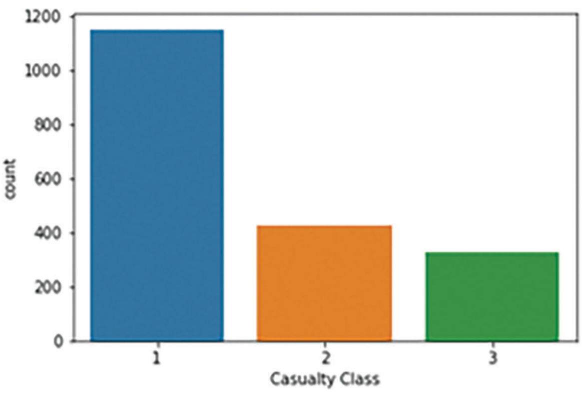

Fig. 4. Distribution of casualty class for the severity prediction for traffic accidents

In this study, we then use the TRAFFIC ACCIDENTS_2019_LEEDS data from the office of Road Safety of the Department of Transport. The classification labels represent each of the data sets. In this database there were 1152 accidents classified as pedestrian, 405 classified as driver or rider, and the remaining 350 accidents were classified as vehicle or pedestrian passenger (see Fig. 4).

Tab. 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.

4.2 Balancing the Database

As shown in the following figure (Fig. 4), the database is unbalanced, because the number of each class is quite different (1 is pedestrian, 2 is vehicle or pillion passenger, 3 is driver or rider).

To balance the database, there are two possibilities: Up sampling, or resampling the values to make their count equal to the class label with the higher count, or Down sampling, picking n samples from each class label where n = number of samples in class with least count.

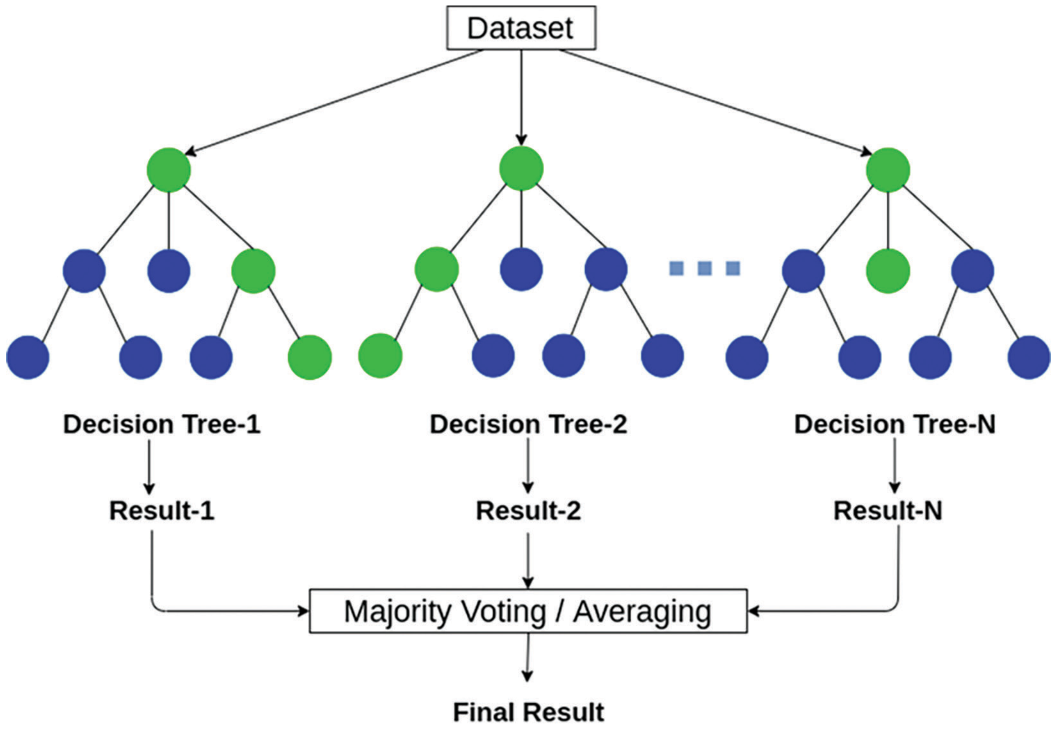

Fig. 5. An overview of random forest

Tab 2. Dataset details after augmentation

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

In this study, we chose to expand the database. We obtained 1152 records for each class, for a total of 3456 records after augmentation (see Table 2).

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 the training set, and 692 features for the test set, in Table 2. This processing was done by capturing in real time all incoming datasets stored in the Hadoop Distributed File System cluster. After this we have called SparkMlLib core to use all the pre-implemented LamBig Data Analytic functions. We took an ANN that consists of an input layer. Fig. 5 shows the random forest algorithm, which combines the output of multiple (randomly created) decision trees to generate the final output.

5. Experimental Results and Discussion

5.1 Evaluation Metrics

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:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Determinants are True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

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

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

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

$$Recal = \frac{TP}{TP + FN} \tag{3}$$

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

F1 score is Precision and Recall in a symbiotic relationship. For unbalanced data, the F1 score is a superior performance statistic than the accuracy metric [37].

$$F_1 = 2 \times \frac{Precsion \times Recall}{Precsion + Recall} \tag{4}$$

5.2 Experimental Setting

We compared the performance of our model to the performance of the ANN, SVM, and RF approaches. This processing is done by capturing in real time all incoming datasets, which are then stored in the Hadoop Distributed File System cluster. After that, we have called upon SparkMlLib core to use all the pre-implemented LamBig Data Analytic functions. We preserved the class ratio between Pedestrian, Vehicle or pillion passenger, and Driver or rider, where the datasets were randomly split into training and test data. Each of the models we tested was trained using training 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 the TRAFFIC ACCIDENTS_2019_LEEDS dataset. The algorithms were created utilizing the Python scikit-learn toolkit and the hyperparameter settings provided.

Fig. 6 represents the confusion matrix for Pedestrian, Vehicle or pillion passenger, or Driver or rider using the 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 several performance measures—precision (PPV), sensitivity or recall, specificity, area under the curve (AUC), F1 score. Fig. 7 also presents the confusion matrix for Pedestrian, Vehicle or pillion passenger, or Driver or rider classification using the random forest model and SVM model. The model that Fig. 8 represents is the Train and validation accuracy curve.

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

6. Discussion

To predict the severity of traffic accidents, we proposed a solution based on big data and machine learning models. This processing was done by capturing in real time all incoming datasets stored in the Hadoop Distributed File System cluster. We then called on SparkMlLib core to use all the pre-implemented LamBig Data Analytic functions. 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. For the 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

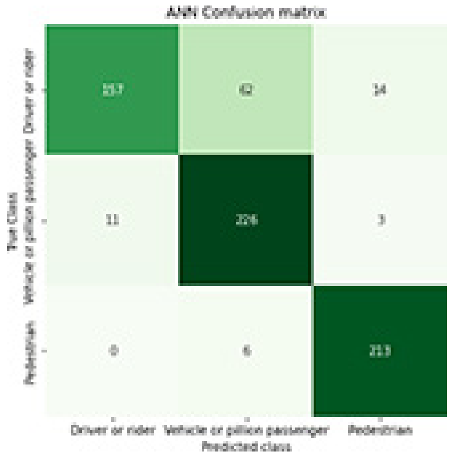

Fig. 6. Confusion matrix for Pedestrian, Vehicle or pillion passenger, or Driver or rider classification using ANN model.

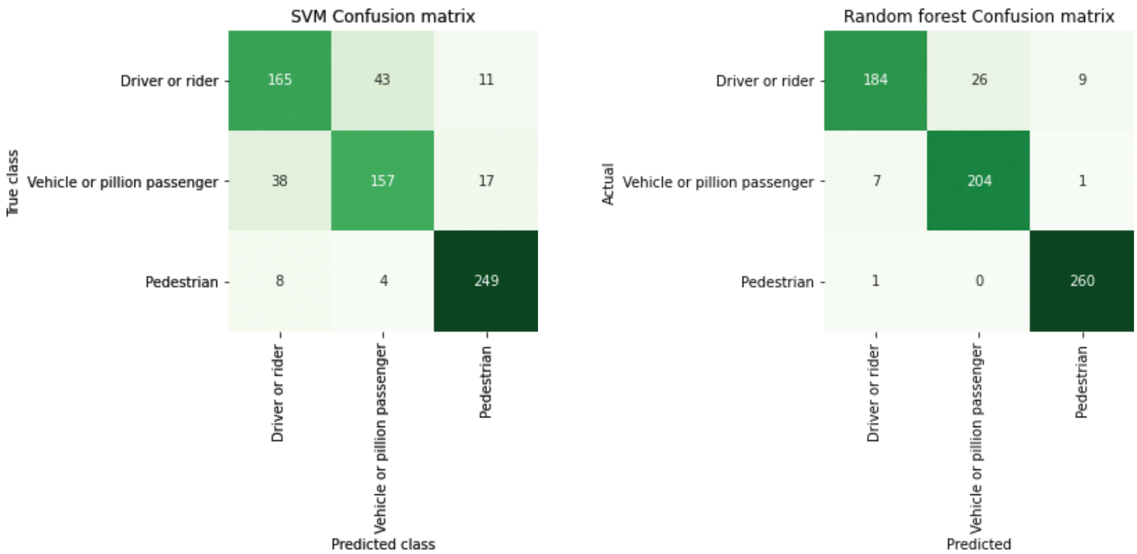

Fig. 7. Confusion matrix for Pedestrian, Vehicle or pillion passenger, or Driver or rider classification using Random Forest model (a) SVM model and (b) Random Forest model.

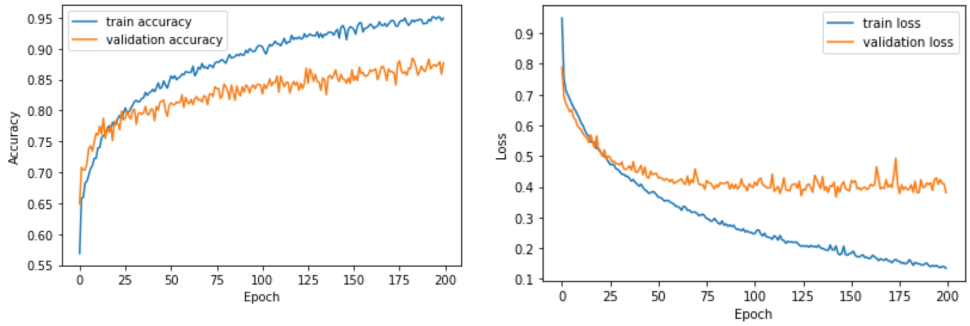

Fig. 8. Train and validation accuracy curve (a) Train and validation accuracy curve, (b) Train and validation loss curve.

Tab. 3. Values obtained for the different metrics

	KNN	Random Forest	SVM	ANN
Accuracy	0.62	0.9364161849710982	0.8251445086705202	0.8771676300578035
precision	0.38	0.9382125952919493	0.8222978546756788	0.8788867858874835
Recall	0.27	0.9364161849710982	0.8251445086705202	0.8771676300578035
F1 score	0.28	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)

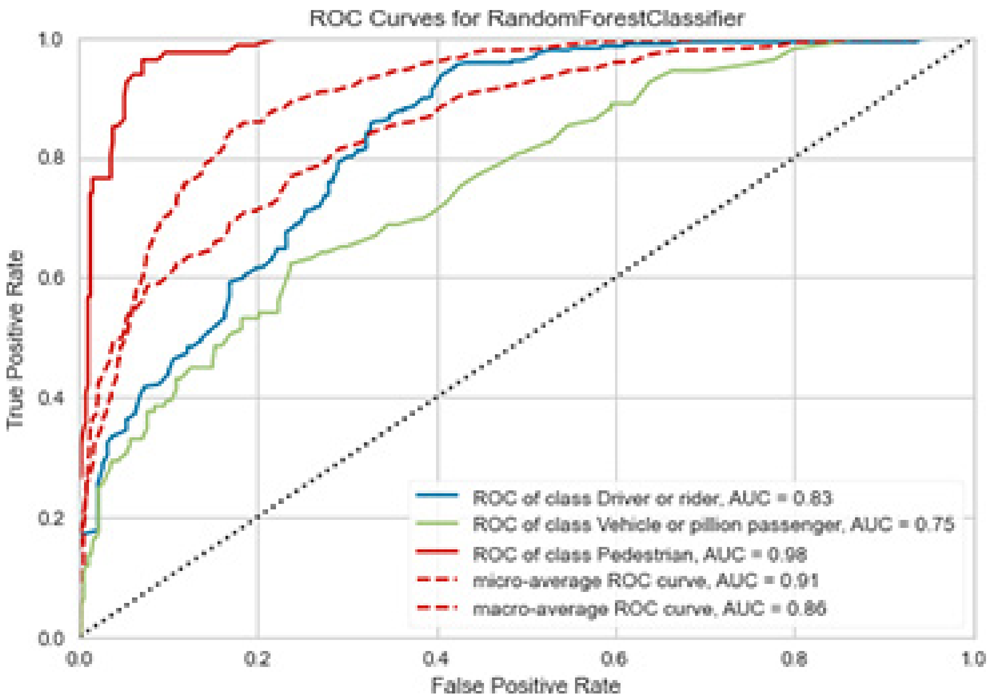

Fig. 9. Classifier ROC curve for Random Forest Classifier.

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. Finally, Fig. 9 represents the receiver operating characteristic (ROC) curves for each class in the random forest model, showing the true positive rate versus false positive rate as the classification threshold is varied between 0 and 1. The ROC curve for each model is an average of 10 curves from the tenfold cross-validation, determined by the trapezoid rule.

7. Conclusion

In this work, we propose a solution based on big data and machine learning models for prediction of traffic accidents. This processing is done by capturing in real time all incoming datasets, which are stored in the Hadoop Distributed File System cluster. Then, we called upon SparkMlLib core to use all the pre-implemented LamBig Data Analytic functions. Next, we focused on severity prediction for traffic accidents, which is a huge step in road accident management. After that, this issue provides important information for emergency logistical transportation. Finally, 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 implemented 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 was 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 will be possible to use more features, and to find best features for classifications for real data in our city. Again, we can extract these selected features from the program file; also, we can implement the cost for the prediction of the gravity of Traffic Accidents. The very important benefit of using the big data paradigm is that it improves the processing of data, and establishes a good rate on road security based on classification and recognition of traffic accidents. We have called directly and in the real time mode on the pre-implemented Machine Learning functions for classifying and predicting the traffic accident in real time.

The new aims and challenge of this work is that we have processed very large data streams in real time mode. This makes possible effective uses of very advanced libraries and a faster system. The obtained results have been tested for accident prevention in different types of areas and roads.

AUTHORS

Imad El Mallahi* – Phd Student in Big Data analytics, traffic accidents, Artificial Intelligence, LISAC Laboratory, Department of Computer Sciences, Faculty of Science, Sidi Mohamed Ben Abdellah University of Fez, Fez, Morocco, imade.elmallahi@usmba.ac.ma.

Jamal Riffi, Hamid Tairi, Mohamed Adnane Mahraz – Sidi Mohammed ben Abdellah University, Faculty of Sciences Dhar el Mahraz, Sidi Mohammed ben Abdellah University, Faculty of Sciences Dhar el Mahraz, LISAC laboratory, Fez, Morocco.

Abderahamane Ez-Zahout – Mohamed V University, Faculty of Sciences, Intelligent Processing Systems & Security Team (IPSS) Computer Science Department, Rabat, Morocco.

*Corresponding author

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., no. 23, 2013, Art. no. 475194.

[3] Shiran, G.; Imaninasab, R.; Khayamim, R. Crash Severity Analysis of Highways Based on Multinomial Logistic Regression Model, Decision Tree Techniques, and Artificial Neural Network: A Modeling Comparison. Sustainability 2021, 13, 5670.

[4] Abdel-Aty, M. Analysis of driver injury severity levels at multiple locations using ordered probit models. J. Saf. Res. 2003, 34, 597–603.

[5] Sze, N.N.;Wong, S.C. Diagnostic analysis of the logistic model for pedestrian injury severity in traffic crashes. Accid. Anal. Prev., vol. 39, 2007, 1267–1278.

[6] Savolainen, P.T.; Mannering, F.; Lord, D.; Quddus, M.A. The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives. Accid. Anal. Prev., vol. 43, 2011, 1666–1676.

[7] Moghaddam, F.R.; Afandizadeh, S.; Ziyadi, M. Prediction of accident severity using artificial neural networks. Int. J. Civ. Eng. vol. 9, 2011, 41–48.

[8] Taamneh, M.; Alkheder, S.; Taamneh, S. Data-mining techniques for traffic accident modeling and prediction in the United Arab Emirates. J. Transp. Saf. Secur. 2017, 9, 146–166. [CrossRef]

[9] Zheng, M.; Li, T.; Zhu, R.; Chen, J.; Ma, Z.F.; Tang, M.J.; Cui, Z.Q.;Wang, Z. Traffic Accident's Severity Prediction: A Deep-Learning Approach-Based CNN Network. IEEE Access 2019, 7, 39897–39910.

[10] Breiman, L. Random forests. Mach. Learn., vol. 45, 2001, 5–32.

[11] Lu, Z.; Long, Z.; Xia, J.; An, C. A Random Forest Model for Travel Mode Identification Based on Mobile Phone Signaling Data. Sustainability, vol. 11, 2019, 5950.

[12] Evans, J.; Waterson, B.; Hamilton, A. Forecasting road traffic conditions using a context-based random forest algorithm. Transp. Plan. Technol., vol 42, 2019, 554–572.

[13] Hamad, K.; Al-Ruzouq, R.; Zeiada,W.; Abu Dabous, S.; Khalil, M.A. Predicting incident duration using random forests. Transp.A-Transp. Sci. vol. 16, 2020, 1269–1293. [CrossRef]

[14] Macioszek, E. Roundabout Entry Capacity Calculation—A Case Study Based on Roundabouts in Tokyo, Japan, and Tokyo Surroundings. Sustainability, vol. 12, 2020, 1533.

[15] Severino, A.; Pappalardo, G.; Curto, S.; Trubia, S.; Olayode, I.O. Safety Evaluation of Flower Roundabout Considering Autonomous Vehicles Operation. Sustainability, vol. 13, 2021, 10120.

[16] Macioszek, E. The Comparison of Models for Critical Headways Estimation at Roundabouts. In Proceedings of the 13th Scientific and Technical Conference on Transport Systems. Theory and Practice (TSTP), Katowice, Poland, 19–21 September, 2016.

[17] 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, 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, 1997, 98113.

[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, 2011, 387399.

[20] S. Al-Ghamdi, Using logistic regression to estimate the influence of accident factors on accident severity, Accident Anal. Prevention, vol. 34, no. 6, 2002. 729741.

[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, 2002, 717727.

[22] K. M. Kockelman and Y. J. Kweon, Driver injury severity: An application of ordered probit models, Accident Anal. Prevention, vol. 34, no. 3, 2002, 313321.

[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, Vol. 29, March, 2014.134–151 https://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, vol. 26, 2000, 32–41.

[29] Yuan F and Cheu RL. Incident detection using support vector machines. Transp Res Part C Emerg Technol. vol. 11, 2003, 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., vol. 59, 2010, 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., vol. 38, 2006, 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., vol. 60, 2013, 142–150.

[34] Lee S-L. Predicting traffic accident severity using classification techniques. Adv Sci Lett., vol. 21, 2015, 3128–3131.

[35] Ez-zahout A., A distributed big data analytics model for people re-identification based dimensionality reduction. International Journal of High Performance Systems Architecture. vol. 10, no. 2, 2021, 57–63.

[36] Haruna Chiroma, Shafi'i M. Abdulhamid, Ibrahim A. T. Hashem, Kayode S. Adewole, Absalom E. Ezugwu, Saidu Abubakar, and Liyana Shuib. Deep Learning-Based Big Data Analytics for Internet of Vehicles: Taxonomy, Challenges, and Research Directions. Hindawi Mathematical Problems in Engineering, vol. 2021, Article ID 9022558, 20 pages, https://doi.org/10.1155/2021/9022558

[37] Asmae Rhanizar, Zineb El Akkaoui, A Predictive Framework of Speed Camera Locations for Road Safety. Computer and Information Science., vol. 12, no. 3, 2019. URL: https://doi.org/10.5539/jmr.v12n3p92

[38] Salma Bouaich, Mahraz, M.A., Riffi, J. et al. Vehicle Counting Based on Road Lines. Pattern Recognit. Image Anal., vol. 31, 2021, 739–748. https://doi.org/10.1134/S1054661821040076

[39] Bouti, A., Mahraz, M.A., Riffi, J. et al. A robust system for road sign detection and classification using LeNet architecture based on convolutional neural network. Soft Comput., vol. 24, 2020, 6721. https://doi.org/10.1007/s00500-019-04307-6

[40] Bouaich, S., Mahraz, M.A., Riffi, J., Tairi, H., Vehicle Detection using Road Lines, 3rd International Conference on Intelligent Computing in Data Sciences, ICDS, 2019, 8942305

[41] Bouaich, S., Mahraz, M.A., Rifii, J., Tairi, H., Vehicle counting system in real-time, 2018 International Conference on Intelligent Systems and Computer Vision, ISCV 2018, May, pp. 1–4