



A Multi-Stream Fusion Model for Accurate Traffic Forecasting

Mrs.A.Anuradha¹, Nandikunta Anusha²

*1 Assistant Professor, Department of CSE, Malla Reddy College of Engineering for Women.,
Maisammaguda., Medchal., TS, India*

*2, B.Tech CSE (20RG1A05A4),
Malla Reddy College of Engineering for Women., Maisammaguda., Medchal., TS, India*

ABSTRACT: Despite more people becoming aware of the dangers of online traffic, deaths caused by vehicle accidents on highways continue to outweigh all other causes. Tragically, road accidents in developing countries are a major threat to both human life and material possessions. There are many potential causes of traffic accidents, and the severity of each accident is dependent on a number of variables. Important components of collapse intensity may be better predicted with the use of data extraction tools. Several features were found to be strongly correlated with the severity of highway accidents in this research that used Random Forest. The primary elements that significantly impact the severity of a surprise are its range, temperature, wind chill, humidity, exposure, and wind directions. This research combines pre-existing deep learning and AI models with RFCNN, or Random Forest and Convolutional Semantic Network, to predict the severity of traffic accidents. It is important to test the suggested approach against many basic learner classifiers to see how well it performs. One set of data included in the research is the United States' accident statistics for the period beginning in February 2016 and ending in June 2020. When predicting the severity of accidents using the 20 most relevant functions, the RFCNN outperformed prior models with a precision of 0.991, an accuracy of 0.974, a recall of 0.986, and an F-score of 0.980.

Keywords: *OSN, Spam, fake account, URL, twitter, social media.*



I INTRODUCTION:

Road traffic accidents are a significant cause of injuries, fatalities, long-term impairments, and property damage. It has an impact on the healthcare system as well as the economy because it puts a strain on the medical facilities. According to data from the Chinese Ministry of Public Security between 2009 and 2011, traffic accidents caused an average of 65123 fatalities and 255540 injuries every year [1]. To lessen the degree of accidental intensity, it is necessary to identify the critical factors that influence how serious a vehicle accident is. Accident Intensity is not a random occurrence; there are patterns that can be predicted and avoided. Unexpected events can Jinjia Zhou, an assistant editor, was responsible for reviewing this submission, accepting it for publication, and recommending that it be evaluated and also avoided. [2] Crash intensity prediction, one of the major issues in crash management, is essential to rescuers in determining the seriousness of website traffic accidents, their potential impact, and in carrying out effective monitoring procedures. A major contributor to injuries, fatalities, lifelong disabilities, and property loss are road traffic accidents. Since it puts a strain on the medical facilities, it has an effect on the health care system in addition to the economy. According to statistics provided by the Chinese Ministry of Public Safety between the years 2009 and 2011, traffic accidents on the internet resulted in an average of 65123 fatalities and 255540 injuries every year (2018). To lower the level of accident severity, it is necessary to identify the critical factors influencing how severe a traffic collision is. Accidental Intensity is not random; it follows predictable patterns that can be recognised and avoided. Unintentional incidents can be evaluated and prevented. Gissane (1965). (1965). Crash seriousness prediction, one of the primary challenges in accident management, has a crucial role to play in helping rescuers assess the seriousness of traffic accidents, their possible effects, and how best to manage accidents. Accidental intensity has been one of the most popular research study locations during the past 20 years. Several statistical methods were being used by researchers for the category of traffic accidents. These techniques aid in determining the cause of traffic accidents. A few examples of traditional statistical-based studies are the combined logit modelling technique Haleem, Alluri, and Gan (2015), the logit model Bedard et al. (2002), and the ordered Probit version Zajac and Ivan (2003). However, Chen and Jovanis note



that these methods are unable to handle multidimensional datasets (2000). Today, machine learning outperforms the conventional analytical techniques in forecasting due to the abundance of available datasets. Sarkar and others (2019). Many scientists have recently concentrated on the study of extent prediction of website traffic crashes. Finding the key factors that significantly affect the severity of online traffic crashes is the main objective of the work of numerous scientists. To carry out such analyses, non-parametric designs, direct models, and data mining techniques have been widely used. Their techniques and also adopted approaches have been analysed in order to describe the literature on severity forecasting of traffic crashes studies.

TARGETS FOR THE TASK

Global website traffic conditions have changed as a result of the sudden increase in traffic on metropolitan streets. Also, it has increased the frequency of traffic accidents, which in the worst cases can result in serious and fatal injuries. Forecasting the severity of accidents is necessary to improve web traffic safety and its management on urban roads. For accident forecasting, many machine learning algorithms are applied. In this study, voting classifiers for the forecast of road crash intensity are compared using tree-based ensemble designs (Random Woodland, AdaBoost, Extra Tree, and Slope Boosting) and an ensemble of two statistical models (Logistic Regression Stochastic Gradient Descent). Random Forest identifies significant characteristics that are highly related to the seriousness of the incident. Study showed that Random Forest, when compared to other methods for classifying the severity of roadway accidents, produced the best classification results, with 0.974 precision, 0.954 precision, 0.930 recall, and a 0.942 F-score using the 20 most important functions.

One of the most serious problems the globe has is traffic accidents, which frequently result in innumerable casualties, injuries, and deaths as well as severe financial hardships. The World Health Organization (WHO) states that India saw 5,18,3626 accidents in 2019. Road accidents, online traffic accidents, negligent drivers, poor road infrastructure, driving in bad weather, and other factors all contribute to these accidents and the damage they cause. In order to identify the severity of injuries, this examination endeavour uses approaches to choose a group of significant variables.



The extent of injuries sustained in traffic accidents can be predicted and estimated using machine learning models.

II RELATED WORK

For individuals, families, and society as a whole, traffic accidents result in enormous economic, financial, and social damages. The estimate of deaths by the World Health Organization (WHO) indicates a disturbing rise in traffic accidents [1]. [2] Each year, more than 50 million people suffer injuries, and over 1.2 million people pass away [1]. Annual increases in the number of traffic accidents around the world are alarming. Every year, traffic-related car accidents claim the lives of almost 1.2 million individuals around the world. A similar pattern has been observed in the nations with the greatest rate of fatal road accidents per 100,000 population, such as Zimbabwe (61.90), Liberia (52.03), Malawi (51.62), Gambia (47.51), Togo (46.62), Tanzania (46.17), Rwanda (45.90), Sao tome (45.52), Burkina Faso (44.94), and Burundi (44.94) [3] The general populace is still dealing with countless catastrophic injuries years after the incident. [4] As a result, worldwide, traffic accidents are now the leading cause of human fatalities and injuries. There are many different causes of accidents. These consist of head-on collisions, single-car accidents, multi-car pileups, rollovers, side-impact crashes, side-swap collisions, hit-and-run occurrences, and drunk driving. The two types of accidents are those that result in significant injury and those that result in minor injury. Any accident injury that requires hospitalization for at least one victim qualifies as a major accident injury. A mild accident injury is one that doesn't necessitate hospitalization for the victim(s). Figure 1 displays a graph of traffic accidents from 2014 to 2019 broken down into three categories: fatalities, injuries, and road wrecks. The Indian Government gave the information (ministry of road transportation as well as motorways study wing, New Delhi). The descriptive version was made using organization learning and clustering methods. [5] [6] Over 140 people have died nationwide as a result of traffic accidents. The states with the most website traffic accidents included Delhi, Maharashtra, Gujarat, Assam, Kerala, Karnataka, Rajasthan, Punjab, and Tamil Nadu. [9] [10] Number 2 displays the numerous online traffic accidents and related injuries. For instance, driving after drinking might result in fatalities or even fractures. Brain injuries, spinal fractures, pelvic cracks, back and spine injuries, head and maxillofacial, rib fractures, fractured



bones, whiplash, cuts and scrapes, internal bleeding, herniated discs, knee trauma, soft tissue injuries, etc. are just a few of the many different types of injuries that can occur. The current method employed less precise equipment learning equations, such as the arbitrary woodland (RF), K-nearest neighbour, and decision tree, to identify accidents, injuries, and damage to cars. [7]. [8] The suggested position is a real-time programme that aids government agencies in reducing the frequency of accidents and the severity of the injuries they cause. Finding unexpected patterns and information in datasets is made possible by an artificial intelligence technique dubbed "not being watched knowing." A method that primarily focuses on information sets rather than training, classification, or testing can quickly detect, analyse, and predict the seriousness of incidents. In order to forecast how bad a traffic collision will be, we employ organization rule mining. [11] [14] Organization rule mining is used to discover the relationships between information elements in large datasets. Association regulation swiftly determines the most significant links between data elements using confidence and support.

Accidents are currently a problem. The host, representative, and environment facilities of traditional epidemiology are briefly reviewed, with a focus on traffic accidents in particular and unexpected injury as a whole. The author is W. Gissane. Some of the crucial aspects of internet traffic crashes are assessed after contrasting the rise in accidents with the drop in contagious diseases over time. Programs aimed at changing behaviour are encouraged not to have any unanticipated short-term benefits. According to this, it would appear that radical advancement is unlikely to occur anytime soon due to financial constraints for environmental changes. The chances for preventing crashes are then considered, however it is indicated that improvements in this area will have the most influence in the future years. According to some, the just emerging field of crash research is an important and unique field of study that needs to be supported.

assessing the seriousness of accidents on the German highway AUTHOR is a topic H. Way and L. Wunsch-Ziegler discuss. We evaluate the amount of accidents that happen on the German Autobahn in the state of North Rhine-Westphalia using data from 2009 to 2011. A multinational logit model, which is separated into four courses—fatal, extreme, mild, and residential property damage—is used to find



statistically significant variables that indicate one of the most serious injuries. In addition, we characterise unobserved diversification using a random criterion version. We investigate the impacts of several factors, including traffic statistics, road conditions, crash kinds, speed limits, the presence of intelligent traffic management systems, the driver's age and sex, and the accident's location. Our findings and studies in different contexts show that accidents happen more frequently and less seriously in the daytime, at interchanges, and on construction sites. Accidents that involve roadside objects, motorcycles and pedestrians, or are the result of blurry eyesight are more likely to be fatal. We assess the measures of the 2011 German traffic safety programme in light of our findings.

EXISTING SYSTEM:

This work employs an ensemble learning method that combines deep learning and machine learning to predict the severity of traffic accidents. The suggested ensemble combines Convolutional Neural Network (CNN) variant RFCNN with Random Forest (RF). The AdaBoost Classifier (AB), Slope Boosting Device (GBM), Random Forest (RF), Bonus Tree (ET), and Vote Classifier artificial intelligence algorithms are the primary learning models used in this investigation. In this experiment, a machine learning voting classifier with two regression-based iterations was used (Logistic Regression and Stochastic Gradient Descent). The US road mishap dataset is applied twice using the suggested architecture RFCNN. The dataset's 48 attributes are first utilized to predict the severity of the accident. We use the RF classifier to calculate the attribute importance value for each function during the initial experimental phase.

PROPOSED SYSTEM:

In order to increase the accuracy and efficacy of category findings, set versions have been widely used. Comparing merged classifiers to individual models, the former can demonstrate significantly higher efficiency. This study use two ensemble versions to forecast the seriousness of a traffic crash in an effort to provide superior results. The first is an ensemble of two machine learning models, and the second is an ensemble of one deep learning and one machine learning design. The suggested method, known as RFCNN voting classifier, integrates RF and CNN while utilising soft voting constraints. The ultimate decision will be based on the class with the highest likelihood.

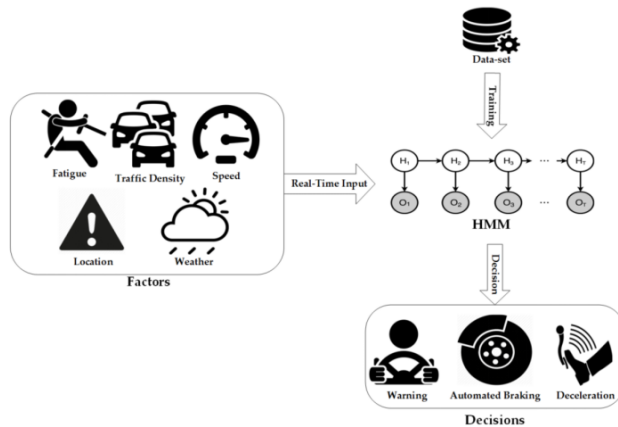


Fig.1. proposed model.

III METHODOLOGY

We conducted the comparison of Tree based ensemble mRoadway accident severity is measured using odels (Random Woodland, AdaBoost, Bonus Tree Classifier, and Slope Boosting Maker) and a collection of regression methods (Voting classifier (LR+SGD)). Also, we used Random Woods to identify 20 significant qualities, or around half of all the dataset's available attributes. In our experiment, we used the dataset's entire set of features as input for the initial phase of all set versions. The most important functions as indicated by RF were used as input for all set designs in the second part of the experiment. The four main findings from the empirical results for unintended intensity prediction are presented. Initially, RF outperformed all previously mentioned set discovery designs to gauge accident seriousness in terms of precision. The highest precision of RF (0.974), using 20 important qualities as input, is achieved in this study. Number 6 contains the precise outcomes of all ensemble designs with all readily available functionalities and important characteristics. Ballot Classifier (LR+SGD) achieved 0.722 precision value when using all functions as input, and it increased to 0.962 accuracy value when using significant features. It is evident that the effectiveness of a collection of regression models' RF has increased when significant functions determined by a tree-based model are used. Figure 6 makes it quite evident that using major functions as input as opposed to all characteristics as input increases the accuracy of all set designs by more than 20%. When compared to the voting classifier, RF achieved a considerably better result with noticeable distinction in terms of accuracy, as shown in figure 7. With 0.784 accuracy value utilising all attributes and 0.954 precision value using key RF-identified



functions, RF achieved higher values. The Voting classifier achieved precision values of 0.692 using all attributes as input and 0.912 using significant features that fall below the RF accuracy level. When compared to vote classifier, tree ensemble versions of RF, AIR CONDITIONING, and other systems get high precision ratings (0.954, 0.922, and 0.928, respectively) (0.912). However, the GBM's precision score is smaller than choosing a classifier that uses all features (0.672) and also uses significant functions (0.902).

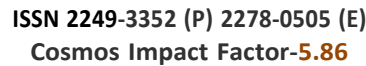
The highest recall value for predicting accident seriousness is 0.93, which RF achieves using significant features to place third in terms of recall (degree of sensitivity). Using significant features, AC and ETC were able to acquire recall values that were virtually identical, 0.901 and 0.904, respectively. By employing all functions, GBM achieved a recall score of 0.741, which was much lower than the classifier's 0.789. However, using significant characteristics, GBM also outperformed the electing classifier with a recall value of 0.921 vs to 0.919. When significant variables found by RF are included in ensemble models, their recall rating is increased.

IV IMPLEMENTATION

To predict crash severity, the author of the suggested research introduces a new machine learning technique called RFCNN (Random Forest Convolution Neural Neural). To increase the accuracy of accident prediction, an algorithmic combination of RF and CNN is proposed. Its effectiveness is then compared to that of other machine learning techniques such as Random Forest (RF), Ada Increase (AIR CONDITIONING), Additional Tree Classifier (ETC), Gradient Boosting (GBM), Ballot Classifier (VC (LR+SGD)), and CNN.

RFCNN is providing high accuracy for all formulas proposed. To train all algorithms, the author used the USA Road Mishap Dataset, which has 47 attributes. The author also evaluated the effectiveness of all formulas using all 47 features (full features) and 23 selected features. RFCNN is providing greater precision for both dataset versions.

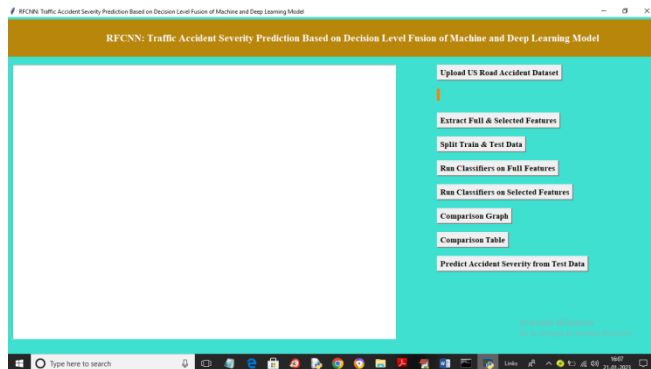
The data that expose dataset details are displayed below.



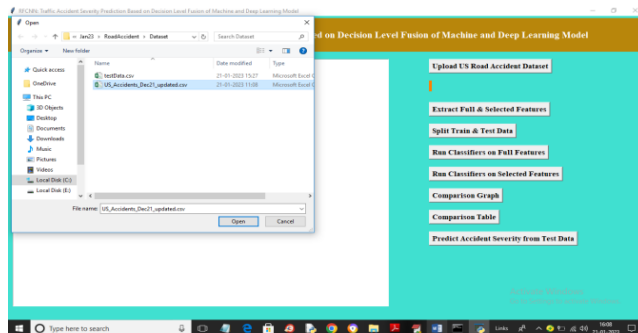
Index in Cosmos
May 2024, Volume 14, ISSUE 2
UGC Approved Journal



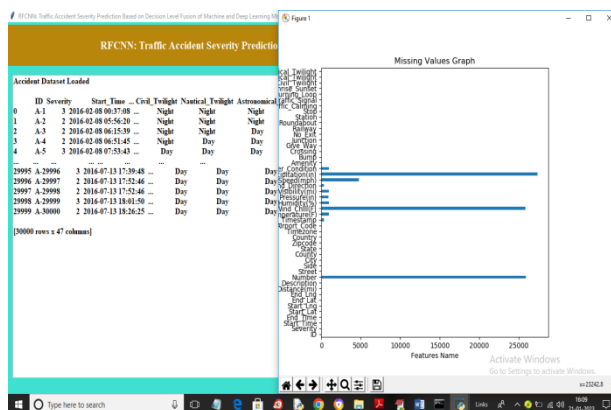
Double-click the 'run.bat' documents to launch the task and see the display below.



To upload the dataset and view the following screen, click the "Upload US Road Accident Dataset" button in the previous screen.



selecting the dataset file in the aforementioned screen, clicking "Open," loading the dataset, and receiving the output shown below



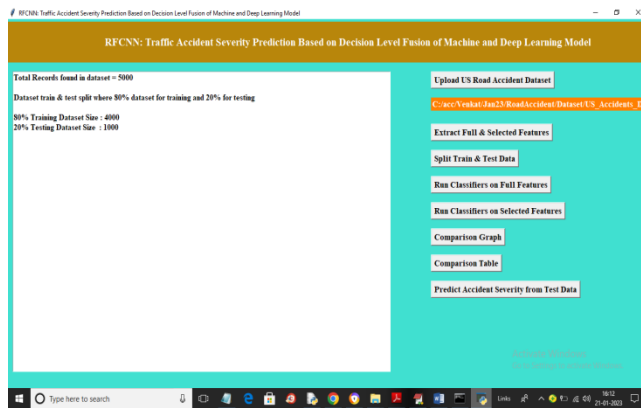
The dataset is loaded at the display above, and the graph's x-axis suggests the quantity of missing values whilst the y-axis lists the names of the features. To update the missing values and separate the total and decided on capabilities datasets, click the "Extract Full & Selected Features" button as visible inside the above graph. This will



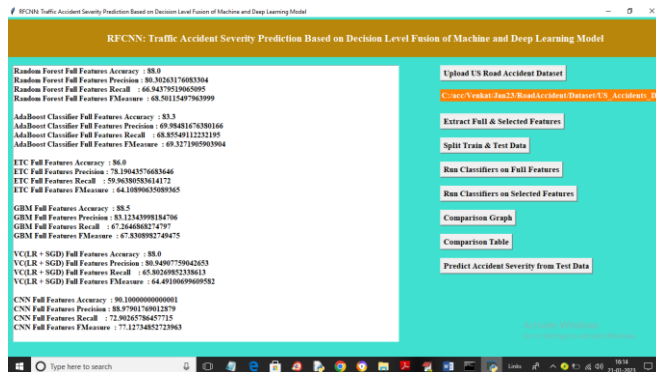
result in the screen that can be seen beneath.



As visible inside the screenshot above, the dataset has forty six complete columns of features. We have selected 23 of those capabilities, and you can see their names in the following lines. Click the "Split Train & Test Data" button to divide the dataset into teach and test, and you may see the end result shown under.



Clicking the "Run Classifiers on Entire Dataset" alternative will educate all device learning classifiers on schooling information and check them on trying out statistics. In the example screen above, we are utilizing 5000 information from the dataset, and the software is the use of four,000 information for training and 1,000 for testing.

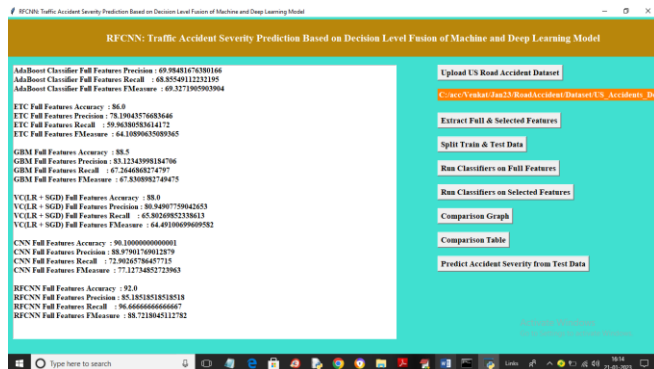




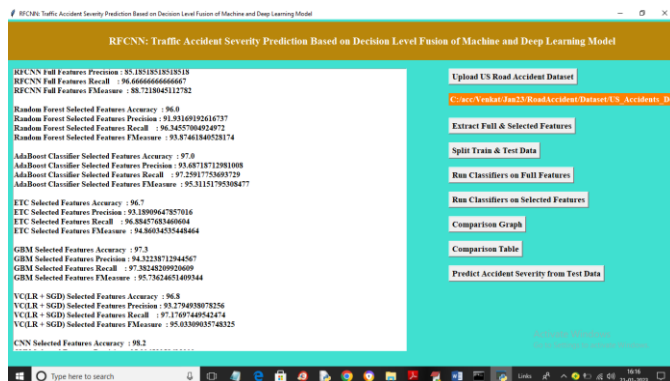
www.ijbar.org

ISSN 2249-3352 (P) 2278-0505 (E)

Cosmos Impact Factor-5.86



The accuracy and different metrics on all features for all algorithms are proven in the two screens above, where we will see that the proposed RFCNN set of rules done a excessive accuracy of 92%. Click the "Run Classifiers in Selected Features" button to educate all algorithms on the selected functions, and the resulting output is shown beneath.



In the display above, every set of rules completed greater than ninety seven% accuracy for some selected functions, even as the RFCNN inspiration carried out a high accuracy of 98.50%. Click the "Comparison Graph" button to view the graph beneath.



We can see that the algorithm with the selected traits carried out properly at the x-axis of the above graph, which lists algorithm names for each complete and selected capabilities. By clicking the "Comparison Table" button, we will reap the output shown under.

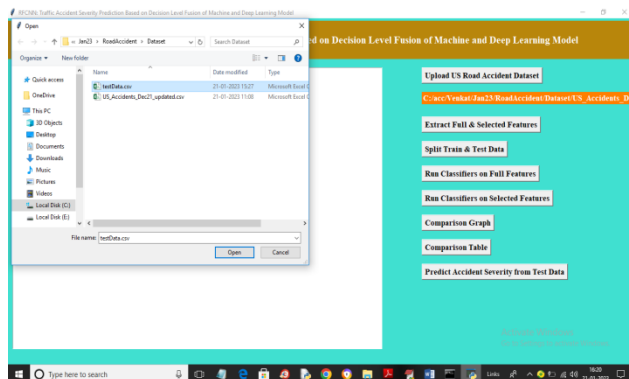
Classification result of all machine learning models using all features

Algorithm Name	Accuracy	Precision	Recall	FSCORE
RF Full Features	88.0	80.30263176083304	66.94379519065093	68.5011497963999
AC Full Features	83.3	69.98441676180166	68.83549112321939	69.3271905903904
ETC Full Features	86.0	70.5040379683666	59.963803814172	64.1089655089565
GRM Full Features	88.3	83.12343995184796	67.254666219797	77.830892748473
VCCLR+SGD Full Features	88.0	80.9490779042615	65.8028981338613	64.4910699606018
CNN Full Features	90.10000000000001	88.97901769012879	72.90265786457171	77.12734852723963
RFCNN Full Features	92.0	85.18518518518518	66.66666666666667	68.728045311782

Classification result of all machine learning models using Selected features

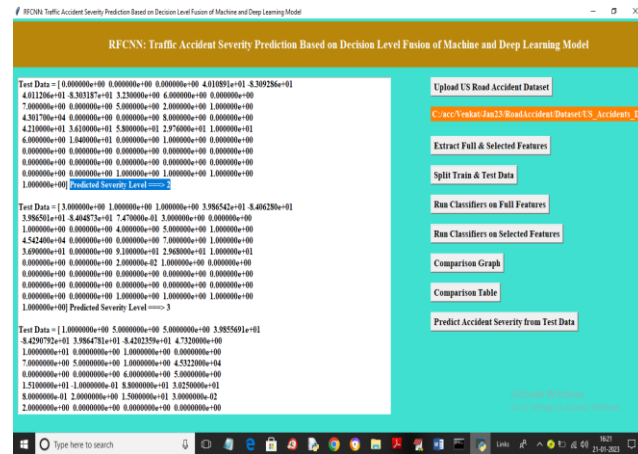
Algorithm Name	Accuracy	Precision	Recall	FSCORE
RF Full Features	96.0	91.8116919261671	96.34351004924872	93.8746380132174
AC Full Features	97.0	93.68718172801008	97.25917733699729	95.31151795308477
ETC Full Features	96.7	93.18909647857016	96.88437683400624	94.86034353448464
GRM Full Features	97.3	94.5228712044567	97.58248209930609	95.73624651409344
VCCLR+SGD Full Features	96.8	93.2794938078256	97.17697449542474	95.0330603748332
CNN Full Features	98.2	95.80658958618909	98.50628420619083	97.12732939254657
RFCNN Full Features	98.5	96.5317243793103	99.37106918238993	97.86435786435787

First table in the above display presentations accuracy and other metrics for all capabilities, at the same time as desk presentations accuracy for only a few features. Both tables display that Propose RFCCN has finished high accuracy. Click the "Predict Accident Severity from Test Data" button to upload take a look at facts and reap the output shown beneath.





Selecting the "TestData.Csv" report within the aforementioned display screen, uploading it, and then clicking "Open" will load the test records and provide the output seen beneath.



The take a look at consequences are shown within the square brackets above the display screen, and the twist of fate severity stages are proven as 2 or 3 after the = arrow image.

V CONCLUSION

It is necessary to control accidents by way of looking at linked elements on the way to increase the effectiveness of the transportation device. In this observe, the RFCNN deep mastering and machine gaining knowledge of model is used to expect the severity level of site visitors injuries. The category effects of RFCNN are greater than those of RF, AC, ETC, GBM, and vote casting classifier (LR+SGD), in line with experimental findings on this studies. Top traits determining unintended severity are distance, temperature, wind Chill, humidity, visibility, and wind direction. Important capabilities are diagnosed with the aid of RF. The most essential features discovered by way of RF are also fed into ensemble fashions, which improve accuracy, precision, keep in mind, and f-score, although RF all over again outperformed with a noticeable distinction. As a end result, it can be concluded that RF is the most effective and green model out of the entire ensemble, and it continually produced correct findings in forecasting accident severity.

VI REFERENCES

[1] World Health Organization, Global Status Report on Road Safety 2015, World Health Org., Geneva, Switzerland, 2015.



- [2] W. Gissane, “Accidents—A modern epidemic,” J. Inst. Health Educ., vol. 3, no. 1, pp. 16–18, 1965.
- [3] H. Manner and L. Wunsch-Ziegler, “Analyzing the severity of accidents on the German autobahn,” Accident Anal. Prevention, vol. 57, pp. 40–48, Aug. 2013.
- [4] J. C. Milton, V. N. Shankar, and F. L. Mannering, “Highway accident severities and the mixed logit model: An exploratory empirical analysis,” Accident Anal. Prevention, vol. 40, no. 1, pp. 260–266, 2008.
- [5] N. V. Malyshkina and F. L. Mannering, “Markov switching multinomial logit model: An application to accident-injury severities,” Accident Anal. Prevention, vol. 41, no. 4, pp. 829–838, Jul. 2009.
- [6] K. Haleem, P. Alluri, and A. Gan, “Analyzing pedestrian crash injury severity at signalized and non-signalized locations,” Accident Anal. Prevention, vol. 81, pp. 14–23, Aug. 2015.
- [7] M. Bedard, G. H. Guyatt, M. J. Stones, and J. P. Hirdes, “The independent contribution of driver, crash, and vehicle characteristics to driver fatalities,” Accident Anal. Prevention, vol. 34, no. 6, pp. 717–727, 2002.
- [8] S. S. Zajac and J. N. Ivan, “Factors influencing injury severity of motor vehicle–crossing pedestrian crashes in rural connecticut,” Accident Anal. Prevention, vol. 35, no. 3, pp. 369–379, May 2003.
- [9] W.-H. Chen and P. P. Jovanis, “Method for identifying factors contributing to driver-injury severity in traffic crashes,” Transp. Res. Rec., J. Transp. Res. Board, vol. 1717, no. 1, pp. 1–9, Jan. 2000.
- [10] S. Sarkar, S. Vinay, R. Raj, J. Maiti, and P. Mitra, “Application of optimized machine learning techniques for prediction of occupational accidents,” Comput. Oper. Res., vol. 106, pp. 210–224, Jun. 2019.
- [11] K. Rumar, “Transport safety visions, targets and strategies: Beyond 2000,” First Eur. Transp. Saf. Lecture, Brussels, Eur. Transp. Saf. Council, pp. 6–8, 1999.
- [12] T. Hummel, “Land use planning in safer transportation network planning,” Leidschendam, Inst. Road Saf. Res., Citeseer, 2001.
- [13] Wadhahi, N.T.S.A., Hussain, S.M., Yosof, K.M., Hussain, S.A., & Singh, A.V., Accidents Detection and Prevention System to reduce Traffic Hazards using IR



Sensors. 2018 7th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), pp. 254-257, 2018.

[14] Zhang, J., Li, Z., Pu, Z., & Xu, C., Comparing Prediction Performance for Crash Injury Severity among Various Machine Learning and Statistical Methods. IEEE Access, pp. 1–1, 2018.

[15] Guang Yu, Shaou lu, Speed and lane changing control optimization research on accident and congestion concentrated section of guang-shen freeway, IEEE Access, pp. 1-8, 2019.

[16] Galatioto, F., Catalano, M., Shaikh, N., McCormick, E., & Johnston, R., Advanced accident prediction models, and impacts assessment. IET Intelligent Transport Systems, 12(9), pp. 1131–1141, 2018.

[17] M., Zheng, et al., Traffic Accident's Severity Prediction: A Deep-Learning ApproachBased CNN Network. IEEE Access, pp. 39897-39910, 2019.

[18] Leijdesdorff, H. A., Gillissen, S., Schipper, I. B., & Krijnen, P., Injury Pattern and Injury Severity of In-Hospital Deceased Road Traffic Accident Victims in The Netherlands: Dutch Road Traffic Accidents Fatalities. World Journal of Surgery, 6, pp. 268-272, 2020.

[19] Zong, F., Chen, X., Tang, J., Yu, P. & Wu, T., Analyzing Traffic Crash Severity With Combination of Information Entropy and Bayesian Network. IEEE Access, pp. 63288– 63302, 2019.

[20] Gao, Z., Pan, R., Yu, R. & Wang, X., Research on Automated Modeling Algorithm Using Association Rules for Traffic Accidents. IEEE International Conference on Big Data and Smart Computing (BigComp), pp. 1-6, 2019.