

Road Accident Prediction System Using Machine Learning Model

Sakibur Rahman Khan
Electrical and Computer
Engineering
North South University
Dhaka, Bangladesh
sakibur.khan@northsouth.edu

Nawar Anbar Fahim
Electrical and Computer
Engineering
North South University
Dhaka, Bangladesh
nawar.fahim@northsouth.edu

Nusrat Zahan Pritha
Electrical and Computer
Engineering
North South University
Dhaka, Bangladesh
nusart.pritha@northsouth.edu

Sayed Bin Harun
Electrical and Computer
Engineering
North South University
Dhaka, Bangladesh
sayed.harun@northsouth.edu

Akhinur Akter Somaya
Electrical and Computer
Engineering
North South University
Dhaka, Bangladesh
akhinur.akter@northsouth.edu

Abstract—Road accidents are one of the most frequent issues in today's transportation sector. They have become a global concern that results in lives, injuries, and other direct and indirect costs. Sadly, car accidents are ranked ninth on the list of the top 10 causes of mortality by the world health organization. It has an impact on many families in addition to our economies. Therefore, a comprehensive analysis is needed to manage this chaotic circumstance. This study extensively used machine learning approaches to examine and anticipate traffic events to determine their severity. We have developed models using two different datasets and a combination of machine learning, and some deep learning methods. One is the US Accident Dataset, which was gathered from Kaggle and included all US Accidental Records from 2016 through 2021. There are 47 attributes and about 4000000 instances in this dataset. Another dataset we used was the Accidental Dataset of Bangladesh (BD Accident Dataset), which was compiled from accident reports from 2000 to 2003 and obtained from the Accident Research Institute (ARI), BUET, Dhaka. This dataset has 20 features and about 460 samples. We were able to forecast the severity of traffic incidents using the open-source US accident dataset and analyze data on the number of crashes by year, by state, on designated days of the week, the ratio of fatal crashes between rural and urban areas, the age of the injured individuals, and the deadliest times to drive. We also made an effort to determine accident severity and analyze data on driver and passenger casualties and the number of accidents by month, year, vehicle type, and district using the BD accident dataset. We focused on anticipating the "traffic accident severity" by incorporating both datasets, and we compared the outcomes between the US and the BD datasets. We removed the feature with too many null values during data pre-processing and filled the other instances with means due to our dataset's large number of cases. A combination of the Synthetic Minority Oversampling Technique (SMOTE) and hyperparameter optimization has been employed. This study developed Decision Trees, Logistic Regression, Random Forest, KNN, ANN, SVM, CNN, and LSTM models to predict accident severity. To enhance the performance of the models, we applied the Voting Classifiers. We also used the explainable AI technique LIME (Local Interpretable Model-agnostic Explanations), which clarifies a machine learning model and makes each prediction comprehensible on its

own. According to the results of the US dataset, CNN received the highest rating of all models, with a training score of 99%, a validation score of 98%, and an F1-Score of 99%. On the contrary, from the BD accidental dataset, the model Random Forest obtained the best training accuracy of 99%, validation accuracy of 94%, and F1-Score of 99%.

Keywords—Road traffic crashes, Forecasting, traffic flow control, accident severity, accident casualties, data analysis, Empirical analysis, driving simulation, Model evaluation.

I. INTRODUCTION

Bangladesh has the densest population in the world, with over 16.14 crore people living in an area of 147577.523 square kilometers [1]. Road accidents are common in a nation with a large population, and the causes of these accidents include aggressive and reckless driving, poor car maintenance, traffic congestion, lax traffic rules, and widespread state usage of the roadways [2]. According to a recent analysis by the Accident Research Institute (ARI) of Bangladesh University of Engineering and Technology (BUET), road crashes cause close to 35,000 injuries and 12,000 fatalities annually on average [3]. According to the World Health Organization's (WHO) 2017 road safety technical package, over 1200 people die as millions more are harmed in automobile accidents globally every year [4]. Road accidents in Bangladesh are increasing drastically every day, which has a considerable negative influence on the economy and society of our country. The data analysis will allow us to find methods for significantly minimizing these incidents. In recent years, scholars have focused their efforts on road accident analysis in an attempt to uncover the characteristics that have a significant impact on automobile accidents.

Most research procedures rely on statistical data or a quick survey via questions or interviews. However, utilizing these antiquated approaches, getting a better and more precise outcome is not easy. The fundamental challenge is that cognitive biases in vehicle crashes are challenging to investigate using conventional research methods. Direct observation is difficult since occurrences are typically unexpected and unplanned. As a result, it is not easy to find entirely correct data. It is crucial to employ a cutting-edge approach capable of producing excellent analytical findings. In this case, machine learning is an even more sophisticated branch of artificial intelligence that can enhance results. Unfortunately, the few studies that