Prediction of Road Accidents Using Machine Learning Algorithm

Nara Bhavyasree[1] , Nara Dhanusree[2]

1. Student, Computer Science and Engineering, Amrita School of Engineering, bhavyasreenara9@gmail.com
2. Student, Computer Science and Engineering, KL University, dhanusreenara26@gmail.com

ABSTRACT

Today, one of the top priorities of governments is road safety. Need to identifying the causes of road accidents has become the primary goal in reducing the damage caused by traffic accidents. Machine learning and data mining concepts are used to recognize the various factors that influence road accidents and their severity. The application uses at Machine Learning algorithms which includes K-Nearest-Neighbor, Decision tree, Random forest, etc make predictions based on various parameters, within all these models KNN gives the best accuracy. This information can be used to analyze future inputs and improve the system's output accuracy. This model can be improved further to send the accident report to the appropriate authorities, such as hospitals, ambulances, and insurance companies, and can thus be very useful in reducing accident fatality rates in the country.

Keywords— *Machine Learning, KNN Algorithm, Severity, Weather Conditions***.**

I. INTRODUCTION

Around the world, road accidents are escalating at an alarming rate. According to the most recent road accident statistics, people who got eternal rest due to road accidents in India reached 3,89,000 in 2021. Speeding, cell phone use, drunken driving, vehicle condition, non-use of seat belts, road conditions like potholes, cracks, and weather conditions are among the causes of fatalities. As reported by the World Health Organization, National Highways and State Highways were responsible for 10% of traffic accidents throughout the world and India contributes 7% of death. In India, a departure from life happens because of road accidents. The Global Status Report generated by the World Health Organization on road Safety claims over 1.35 million people are murdered yearly and every day 4,000 persons died in traffic accidents involving cars, motorcycles, bicycles, buses, pedestrians, or tucks. India ranks first among 199 countries in terms of death in car accidents. In recent years, road accident analysis has garnered a lot of attention from researchers who are trying to figure out what causes accidents. The idea here is to use machine learning and data mining to study and predict accident-prone areas. Maintaining good road quality is thus critical not just for maintaining an effective road network but also for reducing the probability of traffic accidents. However, due to heavy traffic, inclement weather, and expensive labor expenses, maintaining roads regularly is a difficult undertaking. Because road quality must be maintained regularly, a reliable road condition monitoring system is essential for identifying problematic road segments and allocating limited maintenance resources efficiently. At the moment, most road inspections are carried out manually. The manual approach has two drawbacks: fluctuation and coverage. Data accuracy, record preservation, and data analysis are all important aspects of making efficient use of accident reports. There have been a variety of ways to analyze this subject that has been used in this case. The present approaches for preventing accidents in communities have a number of flaws. The collected data will be examined, integrated, and sorted depending on numerous restrictions using the best-suited algorithm. This estimate will aid in the analysis and identification of the fault as well as the causes of the incidents. For this aim, models are designed utilizing Data records from accidents that will help in understanding the characteristics including parameters such as driver behavior, roadway conditions, lighting conditions, and weather conditions, among others. Model is used to find statistically significant characteristics that might forecast the likelihood of crashes and injuries, as well as to perform and decrease risk factors. The fundamental purpose of the road accident prediction system is to

- Design and inaugurate a model that is thrifty, simple, and ply the KNN Algorithm to persuade the severity of accidents.
- To look at older accidents in the vicinity, which could help us determine exactly where more accidents occur.
- Make forecasts based on factors such as weather, pollution, road structure, and so on.

II. SYSTEM ARCHITECTURE

The target is to find the accurate machine learning algorithm for classifying road surface characteristics from datasets. Because of the enormous cost of lives and injuries, traffic accidents have a tremendous impact on society. There has been a surge in interest in recent years in acknowledging more about the elements that determine the severity of a driver's injuries as a result of an automobile accident. Accident investigation is based on accurate and thorough accident records.

Models are built using accident data records and can help comprehend the features of a variety of variables, such as road conditions like cracks and potholes, as well as smooth roads. Light, weather (rainy, sunny, foggy, etc.), road, and so on the machine learning algorithm will be used to train the model. A KNN technique is used to forecast severity by training the model. This method is used to find statistically significant characteristics that might forecast the likelihood of crashes and injuries, as well as to perform and decrease risk factors.

The suggested system includes the following thorough solution procedure:

- Data Acquisition: Extraction and importing of data.
- Data pre-processing: Cleaning of data and feature extraction/selection.
- Machine Learning Training: Decision Tree, Neural Network and Regression Algorithms.
- Model Evaluation: Testing.
- Output: Prediction of severity.

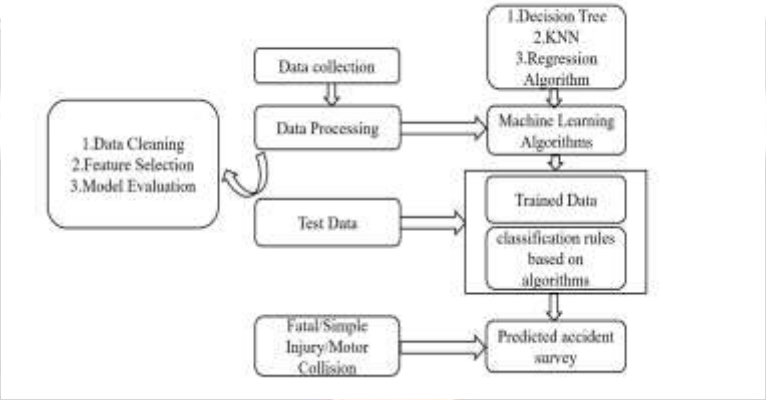

Fig 1:Proposed System Architecture

The methodology of the recommended approaches. The Decision tree is a graphical representation of possible solutions by making decisions under specified circumstances. KNN is a feature similarity-based classification algorithm. It studies the data, analyzes their distances and similarities, and uses K values to divide them into clusters. The distance can be determined in a variety of ways; we'll use the Euclidean distance measurement for this research. Logistic regression is recognized as a statistical technique for modeling the probability of a discrete outcome given an input variable.

A. Data Acquisition

The process of collecting data, including what data is collected, how it is collected, and why it is collected, is known as data acquisition. Data management begins with data acquisition: as soon as the University obtains data, it is responsible for managing it effectively, including adhering to any applicable rules and regulations. The same is true for units and individuals; everyone has a responsibility to manage the data entrusted to their care in an appropriate manner. Acquisition of data includes guidelines such as:

- giving notice of data gathering
- getting consent to data collection
- only collecting particular data
- Obtaining a contract or agreement before collecting data.
- Keep track of data production.

B. Data Pre-processing

Pre-processing of data alludes to the tweaks which are feeding it to the algorithm before feeding. The deed of

embedding data into a clean data set is alluded to as "data pre-processing." In other words, data collected from different sources are raw and need to be processed, without preprocessing making analysis impossible. Pre-processing of data is very much required for accurate results. Data must be appropriately prepared in Machine Learning projects to receive better results from the utilized model. A specific machine learning model requires information in a specified format. A factor to consider is how the data set is organized so that it may also run on numerous machine learning and Deep Learning algorithms simultaneously, considering best algorithm.

C. **Feature Extraction**

A number of features are used in road analysis. Time-domain, frequency domain, and wavelet domain characteristics are the few key categories we look at. Previous research only included a few key criteria to help distinguish between different road conditions. These characteristics, together with the road labels for the dataset, were employed in the training, validation, and testing of models in approaches of machine learning.

D. **Machine Learning Approaches**

It is a part of Artificial Intelligence which involves computers to learn and grow from their experiences without being explicitly instructed. After being trained in machine learning, a computer algorithm can apply what it has learned to similar problems. In this study, we look into a variety of techniques in machine learning to estimate road accidents data obtained from databases.

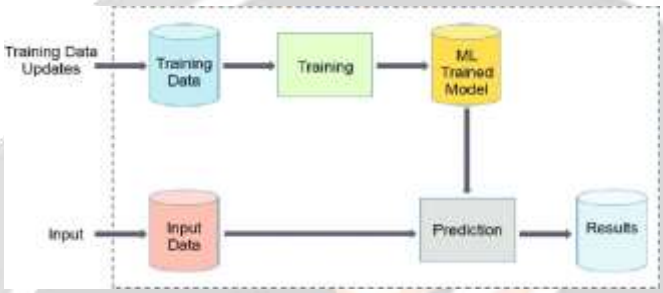

Fig 2:General workflow of ML Algorithm

Fig 2 represents the standard workflow for machine learning techniques such as classification and regression. It starts with a collection of raw data with labels that have already been determined. The input information is processed to extract distinct data properties, referred to as features, that machine learning techniques can exploit. The datastes obtained are trained, validated and testes and are partitioned into three sets after many features and class labels have been extracted. In terms of proportion, the distribution of classes in all three sets is the same. To develop the classifier model and train the algorithm, the training set is used.

KNN Algorithm

The KNN method is one of the machine learning (supervised) methodology that used to solve problems in Classification, prediction stage. Although, it is mostly used in the industry for the categorization of potential issues. The KNN Method Is Effective. 'Feature similarity' is used in the The values of new data points are shown using the KNN machine learning model. Each new data point will be given a value based on how near it identifes the points in the training set. The procedures listed in fig will assist us in comprehending how it works.

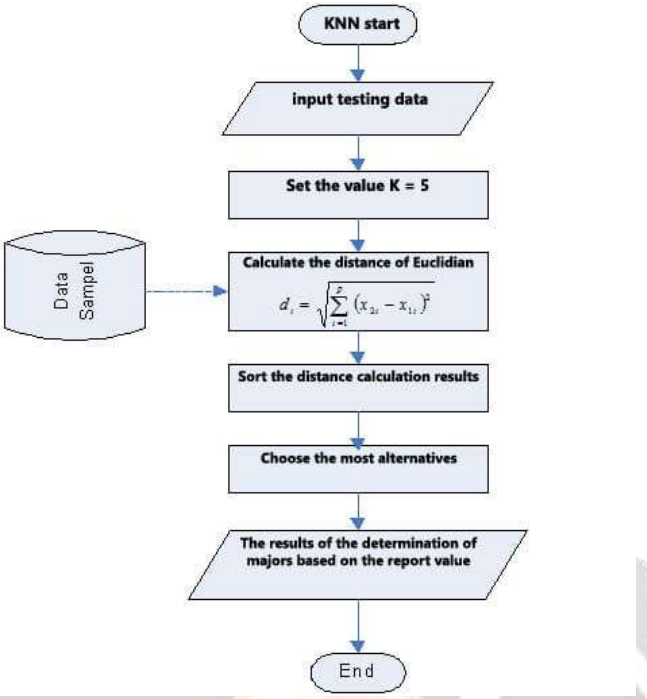

Fig 3: KNN Flowchart

Logistic Regression

Logistic regression one of the popular technique of Machine Learning algorithms which has labled data. From the set of independent factors Based on a categorical dependent variable, logistic regression analyze the model. It gives output as Yes or No, 0 or 1, true or false, and so on, but exact probabilistic values of data records fall between 0 and 1. Types of Logistic Regression: The values of t are given as 0 and 1. Logistic regression is categorized into three groups: Binomial: the dependent variables in logistic regression can only be one of two types in binomial theory. Multinomial: The dependent variable in multinomial logistic regression could be one of three or more unordered sorts, such as "feline," "hound," or "sheep". The dependent variables, such as "low," "mid," or "high" are the three possible ordered types given by the ordinal Logistic regression.

E. Model Evaluation

Various performance evaluation measures for machine learning models are used to assess the performance of all classifiers defined by preceding segment. We consider crucial and important factors for each of the classifiers in order to arrive at a judgment on their performance. Each class label in the data Record represents the total projected class occurrences, while each attribute gives the number of actual class instances. The confusion matrix is used to calculate most classification Metrics based on performance of models for a set of test data.The matrix has two dimensions, predicted values and actual values.

Evaluation Metrics:

Precision: The fraction of all projected observation that is included to the positive class which are positive. The formula for the Evaluation Metric of Precision is as follows:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives + False Positives}}$$

Recall: Recall is the proportion of inspection which are predicted to stay in the positive category but are actually in the positive category. It shows how well the model can recognize a random positive class observation. The formula for the Recall Evaluation Metric is as follows:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives + False Negatives}}$$

F1 Score: F1 Score is a mean Evaluation Metric for calculating a ratio. The harmonic mean of the recall and precision evaluation metrics is called as the F1 Score. In a positive prediction setting, this Evaluation Metric is a measure of model's overall soundness., i.e., how many of the positive inspections in which the model has labelled them as indeed positive.. The Evaluation Metric of F1 Score is given by:

$$\text{F1 Score} = \frac{2*\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

III. RESULTS

The analysis and discussion of collected data, along with the evaluation of each ml prototype capabilities for detecting road accident accuracy, are discussed in this work. The constraints used to measure performance were discussed in preceding section. To predict the accuracy and count of the accident deaths from the data collected based on various parameters are visulaised, and the findings were summarized in Table I. The accuracy of each model used, such as KNN, Logistic Regression, and so on, is shown in these results.

TABLE I : IMPLEMENTATION RESULTS

Models	Accuracy
KNN	89.11
Logistic Regression	88.24
Random Forest	86.84
Decision Tree	78.01

From the above table, we can conclude that KNN machine learning algorithm provides more accurate results than other machine learning algrithms for Road accident analysis.

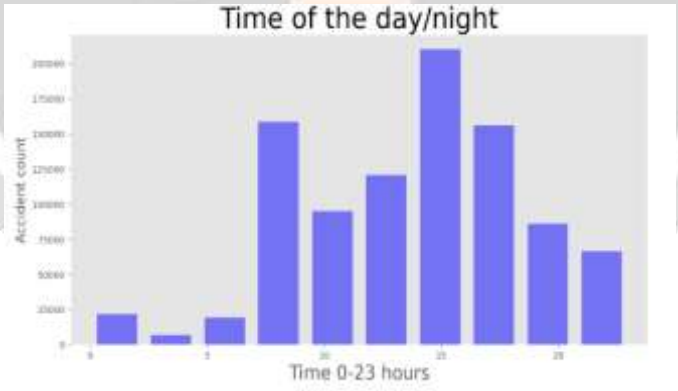

Fig 4: Accident count vs time

From the above fig the count of accidents happening in a day can be visualized and it results in the calculation of accuracy score by various machine learning models.

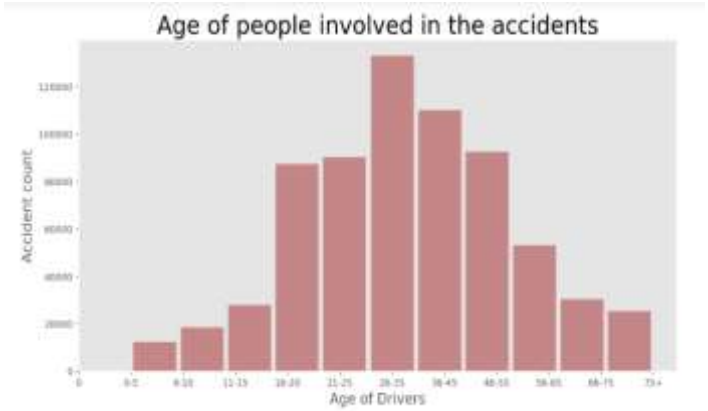

Fig5: Accident count vs driver's age

From the above fig the count of accidents happening in a day can be visualized based on the age of the people involved in accidents .

IV. CONCLUSION

According to the study's findings, ML algorithms were efficient in classifying accident severity based on cracks and potholes, road conditions, weather conditions, and other factors. There exist certain limitations to our current research that are addressed in coming studies. Because our training dataset is so small, precision and accuracy may suffer.

Machine Learning can be used to make an accurate judgment based on prior experience in order to carry out the circumstances from the dataset, the result of the analysis made can be recommended to the required jurisdiction in order to limit the accident numbers. We can recommend the provided techniques to use ML here because of its established and greater accuracy in forecasting road accidents and deaths happening. Furthermore, in order to make it more beneficial, we will strive to develop an adviser system that will be able to predict traffic accidents and inform road users using these techniques. We want to develop a mobile application employing this technology in the future that will provide the user with an accurate prediction while also being incredibly useful and valuable.

V. REFERENCES

[1]. Mubariz Manzoor, Muhammad Umer, Saima Sadiq Saleem Ullah, Hamza Ahmad Madni, abid Ishaq, and Carmen bisogni."Traffic Accident Severity Prediction Based on Decision Level Fusion of Machine and Deep Learning Model." IEEE Access

[2]. N. Sridevi, M.V. Keerthana, Monisha V. Pal, T.R. Nikshitha, P. Jyothi. "Road Accident Analysis Using Machine Learning."International Journal of Research in Engineering, Science and ManagementVolume-3, Issue-5, May-2020 www.ijresm.com|ISSN(Online):2581-5792

[3]. Akanksh Basavaraju, Jing Du, Fujie Zhou, and JimJi. "A Machine Learning Approach to Road Surface Anomaly Assessment Using Smartphone Sensors."IEEE SENSORS JOURNAL, VOL.20, NO.5, MARCH 1, 2020."

[4]. Sahil Dabhade, Sai Mahale, Avinash Chitalkar, Pushkar Gawhad, Vicky Pagare."Road Accident Analysis and Prediction using Machine Learning. "International Journal for Research in Applied Science & Engineering Technology (IJRASET)ISSN: 2321-9653, Volume 8 Issue I, Jan2020

[5]. Vipul Rana, Hemant Joshi, Deepak Parmar, Pradnya Jadhav, Monika Kanojiya. "Road Accident Prediction using Machine Learning Algorithm." International Research Journal of Engineering and Technology Vol: 06 Issue:03 |Mar2019

[6]. Meenu Rani Dey, Utkalika Satapathy, Pranali Bhanse, Bhabendu Kr. Mohanta, Debasish Jena. "Detecting Road Surface Condition using Smartphone Sensors and Machine Learning. 2019 IEEE.

[7]. Jinjun Tang, Lanlan Zheng, Chunyang Han, Weiqi Yin, Yue Zhang, Yajie Zou, Helai Huang "Statistical and machine-learning methods for clearance time prediction of road incidents: A methodology review"Analytic Methods in Accident Research 27, 100123, 2020.

[8]. Jonghak Lee, Taekwan Yoon, Sangil Kwon, Jongtae Lee "Model evaluation for forecasting traffic accident severity in raint seasons using machine learning algorithms:Seoul city study" Applied Sciences 10(1), 2019.

[9]. Lukuman Wahab, Haobin Jiang "A comparative study on machine learning based algorithms for prediction of motorcycle crash severity" PLoS one 14, (4), 2019.

[10]. Nejdet Dogru, Abdulhamit Subasi "Traffic accident detection using random forest classifier" 2018 15th learning and technology conference(L&T), 40-45, 2018.

[11]. J. Patil, M. Prabhu, D. Walavalkar and V.B. Lobo, “Road Accident Analysis Using Machine Learning, ” 2020 IEEE Pune Section International Conference(PuneCon), 2020, pp. 108-112
[12]. M. F. Labib, A. S. Rifat, M. M. Hossain, A. K. Das, and F. Nawrine, "Road Accident Analysis and Prediction of Accident Severity by Using Machine Learning in Bangladesh," 2019 7th International Conference on Smart Computing & Communications (ICSCC), 2019, pp. 1-5.
[13]. S. Sonal and S. Suman, "A Framework for Analysis of Road Accidents," 2018 International Conference on Emerging Trends and Innovations In Engineering And Technological Research (ICETIETR), 2018, pp. 1-5
[14]. D. Al-Dogom, N. Aburaed, M. Al-Saad and S. Almansoori, "Spatio-temporal Analysis and Machine Learning for Traffic Accidents Prediction," 2019