Identify abnormal driving behavior using spatio-temporal analysis

Bhavya Khakhar
AU2140043
bhavya.k1@ahduni.edu.in
B.Tech (CSE)
Ahmedabad University

Harsh Loriya
AU2140154
harsh.l2@ahduni.edu.in
B.Tech (CSE)
Ahmedabad University

Krishna Patel
AU2140170
krishna.p5@ahduni.edu.in
B.Tech (CSE)
Ahmedabad University

Priyam Shah
AU2140187
priyam.s@ahduni.edu.in
B.Tech (CSE)
Ahmedabad University

***Abstract*—The identification of abnormal driving behavior is critical for ensuring road safety and improving traffic management systems. This project aims to develop a binary classifier to identify abnormal driving behavior from trajectory data using spatial and temporal features. The acquisition and labelling of abnormal driving data are, however, difficult, labor-intensive and time-consuming. This situation inspires us to rethink the abnormal driving detection problem. This study investigates the use of machine learning (ML) with spatio temporal analysis for identifying normal and abnormal driving behaviors. By analyzing coordinate data from GPS, we aim to develop models capable of distinguishing between typical driving patterns. We have used Random forest model and their effectiveness is evaluated based on the accuracy and interpretability of the models. This is helpful contribution to enhance road safety and progress in the development of intelligent transportation systems.**

***Index Terms*—Abnormal Driving Behavior, Spatio-temporal Analysis, Trajectory Data, Feature Selection, Random forest algorithm.**

I. Introduction

In recent years, in the era of sensor-equipped vehicles and advances in data analytics, there has been a growing interest in understanding and identifying abnormal driving behaviours. Detecting such behaviours not only enhances road safety, but also facilitates the development of intelligent transportation systems. Traditional approaches often assume uniform driver behaviour within specific road segments, overlooking variations that may indicate abnormal driving patterns based on the certain parameters. This project aims to address this limitation by employing spatio-temporal analysis techniques to discern abnormal driving behaviours from trajectory data. So here, we will distinguish and identify the driving patterns ie. Normal, Abnormal using Random Forest algorithm and using the drone dataset.

II. Methodology

Dataset description and Data Preprocessing: Initially, the data undergoes preprocessing steps such as data cleaning and handling missing values. The provided dataset contains three scenarios, including normal and abnormal behaviours for each scenario. The trajectory dataset was constructed using video data by extracting coordinates from the video. Each file of the dataset contains the following column: frame no. acts as time stamps, left and top columns indicate the bounding box's top and left coordinates for the vehicle in the video, bounding box's width and height are represented by the variables w and h, confidence by conf, and latitude, longitude, and altitude by lat, long, and alt. In the data preprocessing part, we removed certain columns of data from the files that were not needed like conf, lat, long and alt. To understand the data more clearly we added two columns of the centroid coordinates of the bounding box, which were calculated from the left top and width-height given in the dataset.

Data labelling and Splitting: We were given the normal data files and abnormal data files. We have renamed them and moved all files to a single directory by storing old and new filenames in different data file. We have renamed as followed; if it is normal, then "normal_10_1", and for the abnormal file, it is "abnormal_10_1",to make things easy, this format itself gives us the file name and the label of the file in just one argument. After labeling all the files, we have stored them in the corresponding arrays. Then, we performed the train test split and randomly selected the file names. Further, in the training step, we chose the file name randomly, opened each file and stored the required data in an array. Then, trained the model for the whole training data at once. For the testing part, we predicted the labels for each file, not for the whole test data.

$$center_x = left + \frac{w}{2} \qquad center_y = top - \frac{h}{2}$$

The center_x and center_y have been calculated from the bounding box which takes the coordinates of left, top, width and height. The new calculated center coordinates are appended as columns to the data-frame. The center coordinates are used for marking the track lines on the image which verifies the accuracy of the model.

Random Forest Classification: Random Forest is a technique that combines multiple decision trees to make predictions.

- For each tree, a random subset of features is considered for splitting nodes. It is a robust to irrelevant features and noise in data, effectively handling complexity by selecting informative features and makes generalized model.

- This randomness helps reduce over-fitting of the model. The model captures non-linear patterns in data and it

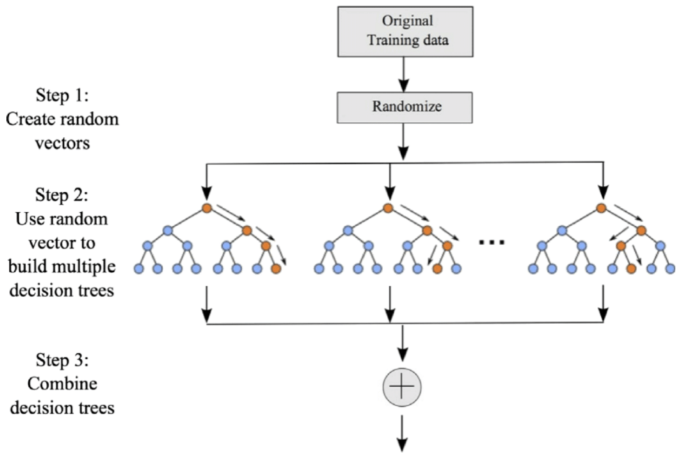

Fig. 1: Random Forest Working.
Source: [https://www.researchgate.net/figure/The-random-forest-algorithm-block-diagram-711_330867263]

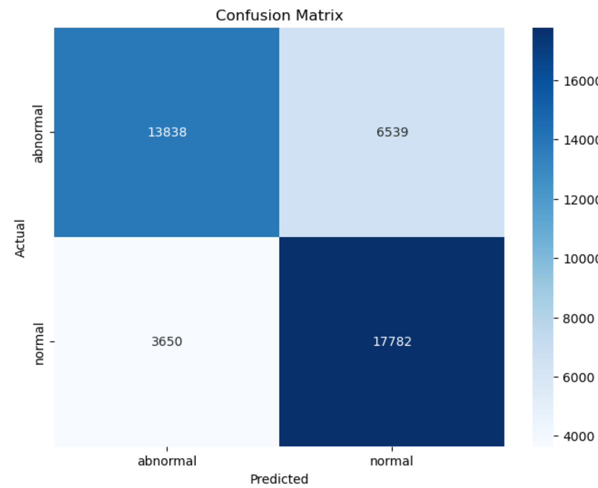

Fig. 2: Random Forest Classifier

can be used to classify driver behavior as normal or abnormal based on majority vote of multiple trees.

- It effectively handles class imbalance in spatio-temporal datasets by averaging multiple decision trees predictions and reducing biases towards majority class which improves the performance.

- For example, abnormal behavior could include sudden acceleration or deceleration, erratic steering, exceeding speed limits, etc. Once the model is trained, it needs to be evaluated to assess its performance. This involves using a separate test dataset to measure metrics such as accuracy, precision, recall, and F1-score.

III. Results

The project is evaluated on real-world trajectory data, which was obtained by breaking video data into separate frames. Experimental results made by feature selection and Random forest classification give an overall accuracy of approximately 78 percent. We make use of visualization techniques to show the expected trajectories and point out unusual patterns. Driving patterns and anomalies can be seen by drawing connecting lines between consecutive locations in trajectories. Images with annotations make results easier to understand and make qualitative analysis easier.

IV. Confusion Matrix

We have made a confusion matrix (See Fig.2) that measures the performance for the model. It helps in understanding the performance of algorithm by providing the analysis of the model's predictions compared to the actual ground truth across different classes. It consists of 4 values ie. True Positive, True Negative, False Positive, and False Negative.

1. True Positive(TP): These are the cases where the model predicted the class correctly, and the actual class is also positive.

2. True Negative(TN): These are the cases where the model predicted the class correctly, and the actual class is also negative.

3. False Positive(FP): These are the cases where the model predicted the class as positive, but the actual class is negative.

4. False Negative(FN): These are the cases where the model predicted the class as negative, but the actual class is positive.

Fig.1 shows that the model correctly identified 13838 normal and abnormal activities, but misclassified 6539 abnormal as normal and 3650 normal as abnormal. This indicates areas for improvement in precision and recall.

$$\textbf{Accuracy:} \frac{TP+TN}{TP+TN+FP+FN}$$

$$\textbf{Precision:} \frac{TP}{TP+FP}$$

$$\textbf{Recall (Sensitivity):} \frac{TP}{TP+FN}$$

$$\textbf{Specificity:} \frac{TN}{TN+FP}$$

$$\textbf{F1-score:} \frac{2\times Precision\times Recall}{Precision+Recall}$$

Based on the above formulas, a Table of Classification Report is formed. See TABLE.

Class	Precision	Recall	F1-score	Support
Abnormal	0.79	0.68	0.73	20,377
Normal	0.73	0.83	0.78	21,432

TABLE I: Classification Report

Fig. 3: Normal_10

V. Failed Results

Fig.2 represents normal driving but it predicts abnormal driving. The path from the blue point (start) to the yellow point (end) seems to follow the correct traffic rules and road markings. This includes proper use of lanes and adherence to traffic signals and signs. The model will predict it as abnormal driving. It may be due to temporary distractions of driver, Speeding in Emergency Situations, Vehicle Malfunctions, etc.

Fig.3 represents abnormal driving but it will predict normal driving. The path here indicates that the vehicles are not adhering to the standard lanes and traffic rules after a yellow point that was marked in the given image. The path from the blue point (start) to the yellow point (end) deviates from the standard lanes, possibly indicating reckless or abnormal driving behavior. This might be due to sudden braking or sudden acceleration. As the model will predict normal driving first, and actually it is abnormal driving, the model will misclassify it.

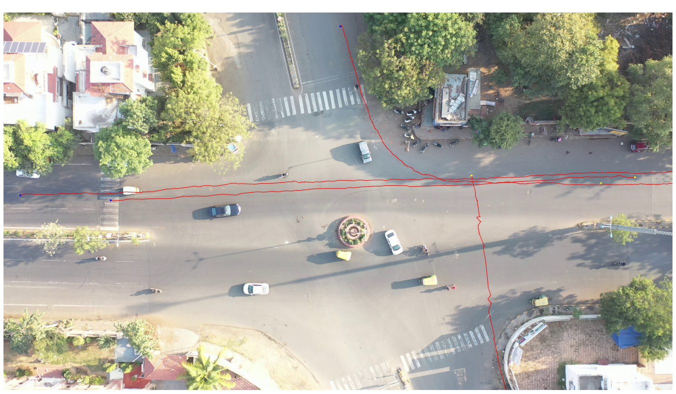

Fig. 4: Abnormal_10

VI. Discussions

A. *Previous Approaches Used*

- Instead of labeling each cell as we did in the previous approach and then dividing the files into two distinct folders, we now label each file, which helps us identify better patterns in the data and increases the accuracy.

- Next, we employed a variety of classification strategies, including Random Forest, Logistic Regression, and Support Vector Machine (SVM). Since random forest provided the highest level of accuracy, we used it.

B. *Performance Evaluation:*

- Our model achieved an accuracy of approximately 78% in distinguishing between normal and abnormal driving trajectories

C. *Analysis of Misclassifications:*

- Normal and Abnormal driving at the same time in a specific path can lead to misclassification.
- For eg. A car is said to be normal when it does not change lanes according to traffic rule but when it takes a sharp turn/suddenly brakes/suddenly accelerates, it is said to be abnormal. But the model will determine this as Normal driving.

- Another reason can be insufficient data. In Fig 2, you can see that on the right, there is also a visible path. Because of insufficient data, the model is unable to classify the behaviour based on that number of data points.

- It is evident that certain driving scenarios may not have been adequately represented in the training data, leading to biases or inconsistencies in the classification process.

The successfull identification of abnormal driving behaviours holds significant implications for various applications, including traffic management, driver assistance systems, and anomaly detection in surveillance scenarios. The use of spatio-temporal analysis techniques coupled with machine learning algorithms enables a nuanced understanding of driver behaviour and enhances the capabilities of intelligent transportation systems.

VII. Conclusion

In conclusion, the domain of road safety is one important area of influence. Abrupt stops, erratic lane changes, and aggressive driving are examples of unusual driving behaviors that often account for a significant portion of traffic fatalities and accidents worldwide. By accurately recognizing and proactively addressing these behaviors, we may lower hazards and improve overall road safety for all users of the road, including bicycles, pedestrians, and passengers. Our project presents a novel approach for identifying abnormal driving behaviours using spatio-temporal analysis and machine learning techniques using random forest model. By leveraging trajectory data and classification using Random forest classifier, the proposed method achieves promising results in distinguishing abnormal driving behaviours from normal ones. Future work may involve further refinement of feature selection strategies and exploration of advanced machine learning algorithms to enhance classification performance.

REFERENCES

[1] Yagnik M Bhavsar, Mazad S Zaveri, Mehul S Raval, Shaheriar B Zaveri, "Vision-based investigation of road traffic and violations at urban roundabout in India using UAV video: A case study," *Transportation Engineering*, vol. 14, p. 100207, 2023. Available: https://doi.org/10.1016/j.treng.2023.100207

[2] Yu Zhang, Yingying He, and Likai Zhang, "Recognition method of abnormal driving behavior using the bidirectional gated recurrent unit and convolutional neural network," *Physica A: Statistical Mechanics and its Applications*, 2022. Available: https://doi.org/10.1016/j.physa.2022.128317

[3] Yongfeng Ma et al., "Real-time detection of abnormal driving behavior based on long short-term memory network and regression residuals," *Transportation Research Part C: Emerging Technologies*, 2022. Available: https://doi.org/10.1016/j.trc.2022.103983

[4] Xiaodi Huang et al., "Abnormal driving behavior detection based on an improved ant colony algorithm," *Applied Artificial Intelligence*, 2023. Available: https://doi.org/10.1080/08839514.2023.2216060

[5] GeeksforGeeks, "Confusion Matrix in Machine Learning," GeeksforGeeks, Available: https://www.geeksforgeeks.org/confusion-matrix-machine-learning/