Aim: The proposed study aims to perform detection of wrong-lane accidents utilizing the Support Vector Machine (SVM) algorithm and compare accuracy with the Random Forest (RF) algorithm. Materials and Methods: Support Vector Machine is applied on a road accident dataset that consists of 1834 records. A machine learning strategy for detecting wrong-lane accidents has been suggested and developed that compares Support Vector Machine with Random Forest. Sample size was calculated as 21 in each group using G power. Sample size was calculated using clinical analysis, with alpha and beta values of 0.05 and 0.5, 95% confidence, 80% pre-test power and enrolment ratio is 1. The accuracy of the detection of wrong-lane accidents was evaluated and recorded. Results: The accuracy was maximum in detection of wrong-lane accidents using Support Vector Machine (89.90%) with minimum mean error when compared with Random Forest (89.77%) and attained significance value of p = 0.02. Conclusion: The study proves that Support Vector Machine Algorithm exhibits better accuracy than Random Forest in detection of wrong-lane accidents.

Keywords: Support Vector Machine, Random Forest, Accident Detection, Wrong-Lane, Data Mining, Classification, Innovative Kernel Based Approach.

DOI: 10.47750/pnr.2022.13.S03.084

INTRODUCTION

The wrong-lane accidents are a common difficulty in the densely populated countries. The capability of the roadway isn't always enough for the developing wide variety of motors and hence imbalance is created (Clarke, Forsyth, and Wright 1998). Every 12 months because of wrong-lane riding accidents, there are numerous deaths and authorities property harm too, which causes massive losses to the countries (Zhao et al. 2021). The drivers do now no longer comply with visitors policies and take gain of riding within the wrong-facet in instances of crimson visitors signals (K.m. and Umamaheswari 2020). It will increase visitors on one facet and hamper visitors greatly. It additionally will increase the opportunity of head-on collision in several instances (Alkhorshid et al. 2016). About 355 humans die each 12 months because of the crashes in wrong-manner riding within the United States. So, it's very essential to prevent drivers from riding on the incorrect facet. To make certain of it, people who don’t comply with visitors' policies want to discover and strict regulation has to be applied (M et al. 2021).

Most referred articles similar to this work have been explored. Around 45 relative articles are published in IEEE Xplore were related to this work in google scholar. (Vasavi 2016) Classification algorithm and Innovative Kernel Based Approach is used widely to improve the detection of accidents. (Patil et al. 2020) proposed Innovative Kernel Based Approach of the system's performance is studied in terms of the classification accuracy, and the results reveal that it has a lot of potential for forecasting the detection of accidents accurately. (Chen and Chen 2020) proposed a criteria based spatial grouping of applications as Innovative Kernel Based Approach with Support Vector Machine performed by other models by achieving 77.3% to predict accidents. (Oza and D. Y Patil School of Engineering Lohegaon 2020) proposed a feature selection algorithm with a classifier same as Innovative Kernel Based Approach for designing a high level intelligent system to predict the accidents taking place.
Pradyumna B, et al.: Awareness of Wrong-Lane Accident Detection using Random Forest Compared with SVM Algorithm with Increased Accuracy

Our team has extensive knowledge and research experience that has translated into high-quality publications (Bhansali et al. 2021; Jayanth et al. 2021; Sudhakar, Ravel, and Perumal 2021; Sathiyanamoorthi et al. 2021; Deepanraj et al. 2021; Raju et al. 2021; Arun Prakash et al. 2020; Kamath et al. 2020; Shanmugam et al. 2021; Rajasekaran et al. 2020; Adhinarayan et al. 2020; Rajesh et al. 2020; Aurtherson et al. 2021). The research gap that is identified from the literature survey is that classification models adopting Random Forest require lots of training data and don’t encode the position and orientation of the object into their predictions. And also, the existing approaches have poor accuracy. The main goal of this study is to implement detection of wrong-lane travelling vehicles and improve the classification accuracy by incorporating Support Vector Machine and comparing the performance with Random Forest.

Materials and Methods

The research study was conducted out at the Department of Computer Science and Engineering, Saveetha School of Engineering. This Study was implemented using Jupyter, and the hardware configuration required is an intel i5 processor, 500 GB HDD, 8GB Ram, and the software configuration required is a windows OS, Jupyter. The work was carried out on 1834 records from an accident severity dataset. The accuracy in detection of wrong-lane accidents is ensured through the evaluation of two groups. A total of 10 iterations were carried out for each group to attain greater accuracy. The Study uses an accident severity dataset downloaded from kaggle (Pote n.d.).

Random Forest (RF) - Group 1

Input: Accident dataset
Output: Accuracy

Step 1: Import and read the dataset.
Step 2: Select the features randomly from the dataset.
Step 3: Generate the RF classifier criterion as a parameter.
Step 4: Gini was used as a parameter value.
Step 5: Construct a decision tree using RF classifiers and predict the result for every sample.
Step 6: Voting was performed for every predicted result.
Step 7: Most voted prediction results were selected as the final outcome.

In this study, the Random Classifier class of the sklearn ensemble library is used. It takes citerian as a parameter. “Gini” is used as the parameter value. The dataset is splitted randomly into training (80%) and testing (20%). It selects samples randomly and the decision trees were collected for every sample to predict the result. Voting was performed for every predicted result and the most voted result was selected as the final result. The algorithm uses a Novel Tree Specific Random Forest Classifier (NTSRF).

Support Vector Machine (SVM) - Group 2

Input: Accident dataset
Output: Accuracy

Step 1: Load the dataset.
Step 2: Split the dataset randomly into training (80%) and testing (20%) dataset.
Step 3: Set the target variable.
Step 4: Generate the SVM classifier based on the training set.
Step 5: Train the classifier using the rbf kernel parameter.
Step 6: Predict the testing set based on the training dataset.
Step 7: Evaluate the classifier.

Step 8: Return Accuracy.

Support Vector Machine (SVM) is a regulation-based machine learning algorithm which can be used for both classification and regression challenges. In this study, to train the SVM the svc class of scikit learn library class was used. The dataset is split randomly into training (80%) and testing (20%) sets. The target variable is selected. Then, the SVM classifier based on the training set is generated. Rbf was used as the value of the kernel parameter. The testing set is predicted based on the training set. The SVM classifier is evaluated and the accuracy is calculated.

Initially, the data set is divided into two groups: the training and test sets. Then, the algorithm is tested on the training and test sets. The training and testing sets are changed 10 times based on the size of the test set. Table 1 shows the comparison between the accuracy of RF and SVM for 10 iterations. The different parameters for the analysis can be calculated as follows:

Accuracy = \[
\frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}}
\] (1)

Statistical Analysis
The statistical software SPSS is used in the research for statistical analysis. Group statistics and independent sample t-tests were performed on the experimental results and the graph was built for two groups with two parameters under study (Ogunbemile, n.d.). The independent variables are Speed Limit, Junction Control. The dependent variables are Accuracy, Prediction.

Results
The proposed Innovative Kernel Based Support Vector Machine and Random Forest have been run at a time for performing detection of wrong-lane accidents. Table 1 shows the accuracy achieved during the evaluation of RF algorithm and SVM models for classification with different iterations. Table 2 exhibits the different parameters of the two groups. Accuracy is calculated for RF and SVM. Two-group analysis shows that SVM has higher accuracy (89.90%) than RF. Table 3 shows the statistical analysis of SVM and RF with different sets of test data. Fig. 1 shows the weekday of vehicle collisions. Fig. 2 shows the policing area and collision severity. Fig. 3 shows the day and month of collision. Fig. 4 shows the speed limit and junction detail. Fig. 5 shows the collision severity. Fig. 6 shows the comparison of mean accuracy of SVM and RF algorithms. The average accuracy of the SVM model appears to be higher than that of the RF model. The performance of the SVM algorithm is far superior to that of the RF algorithm. There is no notable difference between the two groups. Therefore, SVM is better than RF. Statistical analysis of two independent groups shows that SVM has a higher mean accuracy (89.90%). The mean error of SVM is a little less than RF.

Discussion
The work shows that SVM is better than RF at detection of wrong-lane accidents in terms of accuracy. From the experimental results performed in Jupyter, the accuracy of SVM is 89.90%, while RF provides the accuracy of 89.77%. This shows that SVM is better than RF. The different parameters such as TP rate, FP rate are also compared. According to the SPSS plot, the proposed SVM classifier performs better in terms of accuracy (89.90%) than the RF algorithm.

The most important aspect in detection of wrong-lane accidents is accuracy. In the study of, a machine learning-based diagnostic system for wrong-lane detection was proposed using an accident dataset (Jiang, Liu, and Zhang 2007). The study used seven classifier performance metrics such as classification accuracy, specificity, sensitivity, Matthews correlation coefficient, and delay execution, as well as popular machine learning algorithms, three feature selection algorithms (Das et al. 2018).
Based on the above summary, if overall prediction performance is the primary concern, then the integrated method should be selected, in which the RF models with only a few notable variables are identified by SVM or important variables identified by the tree, can be entered to achieve more precision (Gazder, Ahmed, and Shahid 2021). If the focus is on major accident prediction performance, the integrated method, where RF models with only a few significant variables identified by SVM, or major variances identified by the Random Forest, or SVM models with only a few significant variables identified by SVM should be selected for greater sensitivity (Dong, Huang, and Zheng 2015). The accuracy of the SVM classification algorithm depends on the training and testing data set size. In our study, the accuracy appears to be better than RF. However, the average error appears to be higher in our proposed work which should be minimized.

The study outcomes are far superior in both statistical and experimental analysis, there are some few limitations in the work. The accuracy assessment cannot show a far better result on larger data sets. In addition, in RF, the average error seems to be more lesser than SVM. It would be preferable if the average error could be considerably reduced. However, by applying optimization algorithm techniques the work can be improved, to achieve lower mean error as a future work. Prior to classification, feature selection algorithms can be employed to increase the classification accuracy of classifiers. Therefore, thanks to the data mining algorithms, the computation time can be reduced and the accuracy of the classification of classifiers can be improved.

**Conclusion**

The results reveal that the proposed Support Vector Machine outperforms Random Forest in terms of Accuracy. The Proposed Support Vector Machine proved with better accuracy (89.90%) when compared with Random Forest (89.77%).

**DECLARATIONS**

**Conflicts of interests**

No conflicts of interest in this manuscript.

**Author Contributions**

Author PB was involved in data collections, data analysis, algorithm framing, implementation and manuscript writing. Author VN was involved in designing the workflow, guidance, and reviewing the manuscript.

**Acknowledgements**

We thank Saveetha School of Engineering, Saveetha Institute of Medical And Technical Sciences (formerly Saveetha University) for providing facilities and continued assistance to complete this study.

**Funding:** We thank the following organisations for providing financial support that enabled us to complete the study.

1. Mass Datta Developers, Chennai, India.
2. Saveetha University.
4. Saveetha School of Engineering.

**References**

TABLES AND FIGURES

Table 1. Accuracy achieved during the evaluation of RF algorithm and SVM models for classification with different iterations.

<table>
<thead>
<tr>
<th>Iteration No.</th>
<th>ACCURACY</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RF</td>
</tr>
<tr>
<td>1</td>
<td>89.77</td>
</tr>
</tbody>
</table>

ACCURACY
Table 2. Experimental analysis in Jupyter for Accuracy for RF and SVM. SVM provides better Accuracy (89.90%) than RF.

<table>
<thead>
<tr>
<th>MODEL</th>
<th>ACCURACY(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RF</td>
<td>89.77</td>
</tr>
<tr>
<td>SVM</td>
<td>89.90</td>
</tr>
</tbody>
</table>

Table 3. Statistical Analysis of Mean, Standard Deviation and Standard Error Mean and Accuracy of RF and SVM algorithms. There is a statistical difference in accuracy values between the data mining algorithms. SVM had the higher mean accuracy (89.3400%) and RF had mean accuracy of (89.2190%).

<table>
<thead>
<tr>
<th>ACCURACY (Algorithm)</th>
<th>N</th>
<th>Mean</th>
<th>sig</th>
<th>Std. Deviation</th>
<th>Std.Error Mean</th>
<th>95% Confidence Interval for Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Upper</td>
</tr>
<tr>
<td>RF</td>
<td>10</td>
<td>89.2320</td>
<td>.02</td>
<td>.83037</td>
<td>.26259</td>
<td>-.70255</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SVM</td>
<td>10</td>
<td>89.3400</td>
<td>.00</td>
<td>.83327</td>
<td>.22079</td>
<td>-.70466</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Pradyumna B, et al.: Awareness of Wrong-Lane Accident Detection using Random Forest Compared with SVM Algorithm with Increased Accuracy

Fig. 1. Weekday of Collision of Vehicles

Fig. 2. Policing Area and Collision Severity
Pradyumna B, et al.: Awareness of Wrong-Lane Accident Detection using Random Forest Compared with SVM Algorithm with Increased Accuracy

Fig. 3. Day And Month of Collision

Fig. 4. Speed Limit And Junction Detail
Fig. 5. Collision Severity

Fig. 6. Comparison of mean accuracy of RF (89.77%) and SVM (89.90%) algorithms. SVM appears to produce more consistent results with higher accuracy. X-axis: SVM vs RF. Y-axis: Mean Accuracy of ± 1 SD.