ENHANCED TRAFFIC INCIDENT DETECTION USING FACTOR ANALYSIS AND WEIGHTED RANDOM FOREST ALGORITHM

Jayashree R[1] , Trisha C[2] , Anushree N[3]

[1]*UG Scholar, Dept of Information Science and Engineering, Bannari Amman Institute of Technology, Tamil Nadu, India*
[2]*UG Scholar, Dept of Information Science and Engineering, Bannari Amman Institute of Technology, Tamil Nadu, India*
[3]*UG Scholar, Dept of Information Science and Engineering, Bannari Amman Institute of Technology, Tamil Nadu, India*

ABSTRACT

In order to reduce casualties and property damage, efficient and precise traffic incident detection is essential. In order to address the issue of unbalanced event data, this work offers a novel methodology known as FA-WRF (Factor Analysis and Weighted Random Forest). This approach combines dimensionality reduction through factor analysis with classification using weighted random forests, data preparation through the Synthetic Minority Over-sampling Technique (SMOTE), and data preparation with SMOTE. The included feature of severity detection is highlighted in this paper, as is the evaluation of the FA-WRF model using well-established metrics such as detection rate, false alarm rate, classification rate, and area under the receiver operating characteristic curve (AUC). The superiority of the FA-WRF model is illustrated using real-world expressway traffic data characterized by imbalanced incidents through thorough comparisons with various machine learning algorithms such as Support Vector machine, k-nearest neighbors, Logistic Regression, and decision trees. In addition to advancing incident detection, our technique has encouraging prospects for enhancing traffic management procedures and well-informed decision-making procedures in the context of transportation networks.

Keywords - *Traffic incident detection, factor analysis, weighted random forest, unbalanced data, and SMOTE analysis.*

---

1. Introduction:

Effective and precise traffic incident detection is essential in today's society to reduce the number of fatalities and property losses[1]. The likelihood of traffic accidents has significantly increased due to the ever-growing vehicle population and growing road networks. To reduce the negative effects of accidents, including traffic congestion, economic losses, and, most importantly, loss of life, it is essential to

quickly identify and respond to incidents[9]. Therefore, it is crucial to develop cutting-edge methodologies that can quickly and accurately detect incidents and gauge their seriousness.

It is essential to promptly report incidents, such as accidents and road closures, in order to reduce the likelihood of fatalities and sizable property losses. This research reveals a novel approach that combines the powerful methods of Factor Analysis (FA) and Weighted Random Forest (WRF) to completely alter the way incident detection is done. While simultaneously broadening its scope to include severity assessment, a crucial aspect frequently relegated to the background in conventional incident detection paradigms, this comprehensive framework, aptly dubbed FA-WRF[3], serves as a beacon of innovation in addressing the persistent challenge posed by imbalanced event data.

The persistent problem of data imbalance, where one class significantly dominates the other, has a negative impact on incident detection efforts. This imbalance shows up in the context of traffic accidents as an exaggerated representation of typical traffic conditions in contrast to the relatively infrequent occurrences of critical accidents. Machine learning models face significant challenges as a result of this skewed distribution because they perform best when identifying the majority class, which includes urgent incidents that call for immediate action, and perform worse when identifying the minority class[8]. The FA-WRF approach emerges as a potent solution to this conundrum, offering a ground-breaking answer to the fundamental imbalance conundrum as well as the subtle incorporation of incident severity assessment.

The FA-WRF method, at its core, is a testament to the synergy of cutting-edge approaches from the fields of data augmentation, ensemble learning, and dimensionality reduction. Factor Analysis, a flexible statistical method, and Weighted Random Forest, a powerful ensemble classifier, together constitute a paradigm shift in the approach to event detection. By orchestrating the conversion of high-dimensional event variables into their latent factors, factor analysis, a sophisticated method from the history of multivariate statistics, is able to mitigate the effects of dimensionality and improve the effectiveness of the classification process[2]. The Weighted Random Forest technique complements this by presenting a flexible and strong ensemble approach, strengthened by carefully allocating weights to samples depending on their class distribution. By allowing the model to give the underrepresented class more weight than other classes, this weight-based schema resolves the imbalance problem and corrects the skewed performance.

Additionally, the FA-WRF technique goes beyond the limitations of incident detection by including severity evaluation within its scope. Determining the scale and consequence of an incident is equally important for good traffic management and resource allocation, even if incident detection frequently deals with binary categorization. The FA-WRF approach integrates severity evaluation, giving authorities wise insight into incident gradations and enabling them to allocate resources wisely based on the urgency and significance of an occurrence[5]. This paradigm change improves the model's effectiveness by providing stakeholders with a comprehensive understanding of the effects of occurrences, enabling them to develop a more strategic response.

The research projects described in this paper have the potential to change how traffic incident detection and severity assessment work. The FA-WRF approach marks a significant advance in solving the imbalance conundrum that incident detection models have long struggled with[10]. In order to overcome the difficulties posed by skewed class distributions, this method makes use of the skillful dimensionality reduction capabilities of Factor Analysis and the adaptability of Weighted Random Forest to unbalanced data. The creative integration of severity assessment emphasizes the holistic nature of the technique, propelling event detection beyond mere binary judgments into a nuanced comprehension of the gravity of incidents. This study aims to herald in a new era of traffic incident management where precision, equity, and informed decision-making meet to provide safer and more resilient transportation networks[7]. It does this by exploring the intricate mechanics and empirical validation of the FA-WRF approach.

2. Literature Survey

1 Nejdet Dogru et al. (2019) used two machine learning algorithms, random forest and SVM, to develop a model to distinguish accident cases from normal cases. The model was trained on simulated data collected from vehicular ad-hoc networks. The model achieved an accuracy of 91.56%, which is better than the accuracy of the SVM algorithm (88.71%) and the ANN algorithm (90.02%). The authors concluded that the **random forest algorithm is a promising approach for traffic accident detection**. The model developed in this paper can be used to send traffic alerts to drivers in real time, which can help reduce the frequency and severity of traffic accidents.

2 Qiang Shang et al. proposed a hybrid method for traffic incident detection using the Random Forest-Recursive Feature Elimination (RF-RFE) algorithm. They addressed the challenges of an imbalanced and small sample size of traffic incident data. The model outperformed SVM and ANN, achieving a higher accuracy rate. It can send real-time traffic alerts to drivers, potentially reducing the frequency and severity of traffic accidents. The RF-RFE algorithm effectively **identifies relevant features from the data**, enabling the model to focus on critical information for accurate incident detection. This research contributes to the development of advanced systems for enhancing traffic management and accident prevention on the roadways.

3 Tian Xie, Qiang Shang, and Yang Yu (2022) conducted research on automated traffic incident detection, focusing on coping with imbalanced and small datasets. They employed the Random Forest and Random Subspace k-Nearest Neighbor (RSKNN) algorithms to effectively address the challenges of imbalanced data. By using RSKNN with **feature variables as input, the study aimed to**

improve the accuracy and reliability of traffic incident detection, contributing to enhanced road safety and traffic management systems. The research highlights the potential of machine learning techniques for automating incident detection and prevention, even in scenarios with limited or imbalanced data.

4 Huansong Zhang and Yongjun Shen (2022) compared machine learning algorithms for incident detection. They used Decision Tree and Random Forest algorithms and found that Random Forest outperformed the others, making it the **best-performing model for traffic incident detection.** The research emphasizes the importance of temporal-spatial features and highlights Random Forest's effectiveness in enhancing incident detection systems for improved traffic management and road safety. The findings contribute to the advancement of incident detection systems, which can lead to improved traffic management and enhanced road safety. Their research contributes to the ongoing efforts to improve traffic management and road safety through advanced data-driven approaches.

5 Jiawei Wang and Xin Li (2021) employed a combination of Time Series Analysis (TSA) and Machine Learning (ML) techniques to improve the detection rate of traffic-related incidents. The objective was to develop a robust and efficient system that could effectively **identify and classify incidents on roads and highways**. The researchers first utilized Time Series Analysis (TSA) to analyze historical traffic data, which consisted of time-stamped information regarding traffic flow, speed, and other relevant factors. By applying TSA, they could uncover patterns, trends, and seasonality within the data, enabling a better understanding of normal traffic behavior and deviations from the norm. This approach leverages historical data and temporal patterns to achieve more accurate incident identification and classification, ultimately contributing to improved traffic safety and efficient traffic management.

6 Zihan Cheng and Boshi Liu focused on traffic incident detection using a combination of **Factor Analysis and the Decision Tree Algorithm**, with a particular emphasis on analyzing potential objective factors that influence traffic incidents. To begin their work, the researchers applied Factor Analysis to the dataset, which allowed them to identify latent variables or factors that contributed to the occurrence of traffic incidents. Factor Analysis is a statistical technique used to uncover underlying patterns and relationships among observed variables. By extracting these latent factors, the researchers could gain a deeper understanding of the complex interactions between various objective factors and their impact on traffic incidents. This approach enhances incident detection accuracy and provides valuable insights into the objective factors influencing traffic incidents, contributing to safer and more efficient traffic management.

7 **Md. Farhan Labib et al.'s** focus was on traffic incident detection, specifically the classification of accident severity, using Factor Analysis and the application of various Machine Learning algorithms, including Decision Trees (DT), k-Nearest Neighbors (KNN), Naive Bayes, and AdaBoost. To begin the study, the researcher employed Factor Analysis to identify underlying patterns and **correlations within the traffic incident data**. Factor Analysis allowed for the extraction of latent variables that influence accident severity. By understanding these latent factors, a deeper insight into the **complex relationships among various objective features**, such as road conditions, weather, and traffic density, could be obtained. This research contributes to a better understanding of the factors driving accident severity and offers practical applications for enhancing road safety and incident management.

8 Teres Augustine et al. focused on traffic incident detection using Factor Analysis in combination with several Machine Learning algorithms, including Logistic Regression (LR), Random Forest (RF), Decision Trees (DT), k-Nearest Neighbors (KNN), XGBoost, and Support Vector Machines (SVM). Teres Augustine implemented multiple Machine Learning algorithms to build predictive models for traffic incident detection. Logistic Regression is a widely used linear classification algorithm that estimates the probability of a binary outcome, **Random Forest algorithm gives the highest accuracy,** making it suitable for incident classification tasks. Their findings highlighted the superior accuracy of the Random Forest algorithm in this context, showcasing its potential for enhancing road safety and traffic management through effective incident detection.

3. Analysis of Traffic Flow Parameter Variation Characteristics and Construction of Initial Incident Variables:

3.1.Analysis of Traffic Flow Parameter Variation Characteristics:

In the complex realm of traffic dynamics, events cause fluctuations in traffic flow parameters and halt the continuous flow of moving traffic. A crucial first step in creating successful incident variables is comprehending these changes and drawing insightful conclusions from them. This part begins a thorough investigation of the complex interactions between incident occurrences and traffic flow factors[6]. Traffic flow parameters are complex measures that include elements like volume, occupancy, and speed. These characteristics display a certain pattern that is characterized by consistency and predictability under regular traffic situations. However, these measures experience observable changes from their baseline values following incidents, such as accidents, road closures, or bad weather. For instance, traffic volume fluctuates, speed declines dramatically, and vehicle occupancy rises as a result of congestion.

Importantly, the specifics of these parameter fluctuations vary according to the type and seriousness of the incident. Minor occurrences could result in a mild decline in speed and a slight rise in occupancy, whereas catastrophic accidents could cause a significant drop in speed, a rise in vehicle occupancy, and possibly even sporadic road closures. Therefore, by carefully analyzing these variances, event-specific patterns that can be used for successful incident detection can be discovered. Utilizing a variety of data sources, aspects of traffic flow parameter variation are analyzed[4]. A wealth of knowledge on traffic flow dynamics is available via real-time sensor data from road cameras, loop detectors, and other sensor networks. To find complex patterns in large datasets, machine learning techniques are frequently used. For instance, clustering algorithms can identify discrete clusters that correlate to different levels of event severity by grouping comparable traffic flow patterns.

Additionally, temporal patterns in parameter fluctuations are captured using time-series analytic approaches. The commencement and progression of incidents can be determined by reviewing data collected across a range of time frames, from seconds to minutes to hours. This temporal granularity is crucial because it makes it possible to spot sudden changes that signal accidents as they happen. Another crucial component of this study is the incorporation of domain knowledge[1]. Deciphering the significance of parameter fluctuations requires the expertise of experts in the field of transportation engineering. They aid in separating anomalies brought on by incidents from those brought on by common occurrences like traffic signal changes or lane closures for repairs.By combining domain knowledge and data-driven analysis, we can better understand how changes in traffic parameters affect incident detection. The foundation for creating useful event variables is laid by an investigation of the characteristics of variation in the traffic flow parameter. Incident-specific signals can be determined by recognizing and measuring the precise changes in speed, occupancy, volume, and other pertinent factors. The foundation of event variables that may be used to train and evaluate machine learning models is made up of these signatures, which are distinguished by their uniqueness and consistency across comparable situations.

An essential component in the process of incident identification and severity evaluation is the examination of traffic flow parameter variation characteristics[3]. The foundation for developing incident variables that are sensitive and particular to distinct incident kinds and severities is set by closely examining the variances in traffic flow metrics during events and figuring out their patterns. This detailed understanding, which is further backed by data-driven insights and subject-matter experience, strengthens the subsequent steps of this technique. The management and reaction strategies for traffic incidents are then improved using this framework.

3.2. Construction of Initial Input Incident Variables

The conversion of subtle variations in traffic flow parameters into pertinent and reliable incident variables is necessary for the accurate detection of traffic events[5]. This section explores the complexities of creating these variables and clarifies three key tactics that work together to produce an extensive incident representation.

1. Combining measured and predicted traffic flow parameters

A reliable method of creating incident variables is the marriage of real-time measurable data with anticipated traffic flow factors. Measured metrics are obtained directly from camera and sensor networks, providing accurate information from the ground. These measurements could, however, show noise, missing data, or differences as a result of sensor errors. Predictive models offer estimates of these parameters with increased accuracy because they frequently draw on past data and machine learning algorithms.

Both sources work together to create a symbiotic relationship. The measured parameters ground the incident variables in reality, and the projected parameters properly and thoroughly depict the event variables[2]. This fusion provides protection against the limitations and errors present in each separate data source. Additionally, it strengthens the incident variables' resistance to situations where sensors might be harmed or where data quality might change.

2. Combining several traffic flow parameters obtained from the same detector:

The various aspects of traffic flow parameters include measurements like speed, occupancy, and volume. Multiple parameters from the same detector are included, enriching the granularity and comprehensiveness of the incident variable as opposed to depending exclusively on one parameter. For instance, in the event of an accident, occupancy may increase as a result of congestion, even though speed may drastically decrease. This varied combination aids in precise detection and offers a more comprehensive knowledge of occurrences.

Additionally, various factors may respond to situations in different ways. For example, a lane closure could not have a significant impact on speed but might immediately boost occupancy. The event variables learn to distinguish between distinct incident circumstances and adapt to varied contexts by recording these complex interactions through a variety of factors.

3. Combining the Same Traffic Flow Parameters from Neighboring Detectors:

Incident detection becomes even more challenging due to the spatial dimension. The field of view of a single detector may not contain an incident; incidents may span several detectors along a route section. A comprehensive picture of the incident's spatial extent and evolution benefits from the inclusion of data from nearby detectors[4]. For instance, a congestion-causing occurrence may begin at one detector and spread to nearby detectors, resulting in a cascade effect.

Incident variables are able to capture the spatiotemporal dynamics of occurrences by averaging the same traffic flow data from nearby detectors. For large-scale accidents, road closures, or incidents involving a series of events, this spatial awareness is especially important[9]. This strategy's full incident representation provides a more precise assessment of event severity and makes it easier for traffic management authorities to respond in a more focused manner.

Essentially, the creation of initial input incident variables is a deliberate and complex procedure that makes use of a combination of predicted and measured parameters, numerous parameters from a single detector, and information from nearby detectors. This clever combination adds precision, detail, and geographical awareness to the incident factors[6]. In the later stages of the FA-WRF approach, factor analysis and weighted random forest are used to detect occurrences and precisely gauge their severity. These variables serve as the system's cornerstone.

The extraction of incident variables by factor analysis and the use of the weighted random forest algorithm for incident identification will be further discussed in the sections that follow. The holistic approach of FA-WRF is revealed through these thorough processes, demonstrating its effectiveness in resolving imbalanced data and revolutionizing the field of traffic incident identification and severity assessment.

4. Expressway Traffic Incident Detection Based on FA-WRF:

A. Motorway Traffic Incident Detection Using Input Variable Extraction Based on FA:

The efficacy of incident detection hinges on the precision of input variables, which aptly capture the nuances of traffic dynamics during incidents. This section delves deeply into the utilization of factor analysis (FA) to extract these pivotal incident variables. Factor analysis, a robust technique for dimensionality reduction, unveils latent components inherent in the observed data[10]. These components provide insights into the intrinsic structure of traffic events and guide the construction of incident variables.

4.1. Let's walk through the intricate steps of this process:

Data Collection:

The basis is a large dataset containing a wide range of traffic flow characteristics from multiple sensors along motorways. These factors, which also include traffic volume (*V*), speed (*S*), and occupancy (*O*), all play a part in the complex web of traffic dynamics[8].

Factor Extraction:

Factor analysis reveals latent components that are responsible for the parameters' observed variances. This can be mathematically stated as:

X = AF + E

Where:

X is the parameter observation matrix.

The matrix of loadings that maps parameters to latent factors is represented by A.

F stands for the latent factor matrix.

E stands for the error term matrix.

Loadings (λ) measure how each parameter affects the latent variables. The loading is represented as λ<sub>ij/sub> for the latent factor j and parameter i.

Identification of Relevant Parameters:

By examining loadings, relevant parameters that significantly contribute to particular latent variables are found. In incident detection, parameters with substantial loadings on certain factor anaspects are important.

Creation of Incident Variables:

Incident variables are created by averaging parameters according to how they load on latent variables[9]. The mathematical expression of the aggregation for incident variable k and parameter i is:

IVsub>k/sub>> = ∑ (λ_{ik/sub> Xsub>i/sub>i})

Where:

- IV_k incident variable k.

- λ_{ik} the loading of parameter i on latent factor k.

- X_i is the parameter i.

Latent components that effectively capture the fluctuations in traffic flow data are retrieved using factor analysis. These variables are the foundation for creating event variables that accurately and comprehensively represent the essence of situations[1]. These composite incident variables overcome the limitations of individual factors through careful aggregation led by loadings, resulting in a refined representation.

	features	Fscore	Pvalues
14	Light_conditions	16.082824	0.000322
20	Age_band_of_casualty	13.778413	0.001019
16	Type_of_collision	10.096323	0.006421
1	Age_band_of_driver	8.915392	0.011589
12	Road_surface_type	6.994806	0.030276
4	Vehicle_driver_relation	5.345345	0.069067
5	Driving_experience	4.499679	0.105416
8	Area_accident_occured	3.616540	0.163937
9	Lanes_or_Medians	3.281615	0.193824
18	Casualty_class	3.216860	0.200202
23	Cause_of_accident	3.193666	0.202537
11	Types_of_Junction	3.086487	0.213687
17	Vehicle_movement	2.200712	0.332753
15	Weather_conditions	1.149345	0.562889
7	Owner_of_vehicle	1.104262	0.575722
6	Type_of_vehicle	1.077671	0.583427

Fig - 1: f score

Overall, the procedure proceeds in a methodical evolution from uninformed event variables to insightful incident variables that are ready to drive the remaining stages of the FA-WRF approach[4]. The model's ability to identify and classify incidents in the midst of complicated traffic dynamics is supported by this careful input variable extraction method.

4.2. WRF-Based Motorway Traffic Incident Detection

4.2.1.The first Random Forest (RF) principle is:

Understanding the foundation of the latter part, the random forest method, is crucial before digging into the specifics of the FA-WRF technique. During training, numerous decision trees are built as part of the random forest ensemble learning technique. The forest becomes more diverse because each tree is formed using a different portion of the data and a different subset of the characteristics. Each tree casts a vote on the final classification before providing a prediction, and the class with the most votes is selected as the prediction. By eliminating overfitting, this method improves generalization.

4.2.2. WRF-based design of an AID algorithm:

The process of building individual decision trees in the random forest starts with bootstrapping, which involves selecting random subsets of the data and replacing them. By introducing improvements to this procedure, the FA-WRF technique may increase the diversity of the trees[6]. This diversity leads to a stronger ability to collect various occurrence patterns, which enhances the effectiveness of detection.

4.2.3. Weighted Random Forest Model:

The FA-WRF method is novel in that weights are added to the random forest algorithm. By reflecting the importance of each instance, these weights make up for the data imbalance caused by incidents being of a minority class. By giving instances of the minority class more weights, weighted random forest ensure that they have a significant impact on the decision-making process. The model can effectively learn from instances of both the majority and minority classes thanks to its dynamic weighting technique.

4.2.4.WRF-Based AID for Motorway Traffic Incidents Working Steps:

The FA-WRF algorithm uses a structured approach to more accurately identify traffic accidents[3]. As was previously said, it starts with the factor analysis-based extraction of event variables. The weighted random forest model receives these incident variables as input, which represent latent incident-related patterns. The weighted random forest adapts its internal structure as it learns from the event data during the training phase to recognize the complex connections between incident variables and incident occurrences[2]. The weighting technique makes sure that the algorithm gives the majority and minority classes the attention they deserve, allowing it to properly record complex occurrence patterns.

The trained FA-WRF model uses the incident variables produced from the ongoing traffic flow parameters as input when it is used for real-time event identification. The model determines whether or not there was an incident by taking into account the group votes of the ensemble of trees. Even in the presence of unbalanced data, the algorithm's decision-making is guided by the weighting approach used during training, ensuring reliable incident detection.

The FA-WRF technique creates a powerful system for detecting motorway traffic incidents by combining the power of weighted random forests with the insights from factor analysis. By utilizing the power of machine learning, the methodology gracefully navigates the difficulties presented by imbalanced data[5].

This method provides a novel and practical technique to improve the accuracy of traffic incident detection and severity evaluation by carefully collecting event variables and training the algorithm to adjust for data imbalance.

4.2.5. AID Algorithm Evaluation Index:

Robust evaluation measures are necessary to quantify the performance of the FA-WRF model, as they improve the accuracy of incident identification and severity assessment[8]. This section goes into a broad range of evaluation indices that, taken together, shed light on the model's effectiveness in spotting incidents, assessing their seriousness, and making judgments.

A crucial metric that quantifies the proportion of correctly classified cases among all examples examined is the accuracy rate. In mathematics, it is written as.

Accuracy Rate = Number of Correctly Classified Instances / Total Number of Instances * 100

The model's ability to distinguish between typical traffic circumstances and events is indicated by its accuracy rate. Although it might seem obvious, accuracy can be deceptive, especially when working with unbalanced datasets. When one class (for example, usual traffic) significantly outweighs the other (incidents), the model may predict the majority class with a high degree of accuracy. Since accuracy only gives a broad picture of performance, it must be combined with other metrics that take into account circumstances involving imbalanced data.

4.2.6. F-Measure:

The F-measure, also known as the F1 score, strikes a compromise between precision and recall, making it an appropriate option for assessing unbalanced datasets[9]. Recall estimates the percentage of true positive predictions among all actual positive instances, while precision measures the percentage of true positive forecasts across all positive predictions. What is meant by the F-measure?

F-measure = (2 * Precision * Recall) / (Precision + Recall)

Models that simultaneously attain high precision and recall are rewarded by the F-measure[8]. This is essential for incident detection since both false positives and false negatives—misclassifying routine traffic as events or failing to identify actual occurrences—have detrimental effects. The F-measure gives a more thorough assessment of the model's performance by striking a balance between these factors, particularly when dealing with imbalanced classes.

4.2.7. Matthews Correlation Coefficient (MCC):

When working with unbalanced datasets, the Matthews Correlation Coefficient (MCC) is a useful indicator for assessing the effectiveness of classification algorithms[6]. It provides a comprehensive assessment of a model's prediction ability by accounting for true positives, true negatives, false positives, and false negatives. The following is the MCC formula:

MCC = (TP *TN - FP *FN) / (sqrt (TP + FP) (TP + FN) (TN + FP) (TN + FN)

In this context, TP stands for true positives, TN for true negatives, FP for false positives, and FN for false negatives. MCC produces a result in the range of -1 and +1, where +1 denotes a perfect prediction, 0 denotes a random forecast, and -1 denotes a total discrepancy between the prediction and the actual result. Due to its consideration of both the actual positive rate and the true negative rate, MCC is particularly useful in situations when the class distribution is unbalanced.

4.2.8. ROC Curve and AUC Value:

When assessing the performance of classification models, particularly when working with different threshold values, ROC curves and the Area Under the Curve (AUC) value are effective tools. The genuine positive rate (sensitivity) versus the false positive rate (1-specificity) when the decision threshold increases are depicted graphically by the ROC curve. The model's capacity to distinguish between classes at various threshold values is summarized by the area under the ROC curve, or AUC[8].

While an AUC score around 0.5 shows random performance, one closer to 1 indicates greater model performance. When working with imbalanced datasets, the ROC curve and AUC are very helpful since they reveal how well the model can distinguish between incidents and non-incidents at various operating points. The threshold's flexibility enables the model's sensitivity and specificity to be fine-tuned in accordance with the demands of incident detection and severity evaluation[10].

These evaluation indices—accuracy rate, F-measure, MCC, ROC curve, and AUC value—provide a thorough evaluation of the performance of the FA-WRF model.

Finally, the careful selection of evaluation indices highlights the seriousness and breadth of this study. These measures demonstrate the FA-WRF model's ability to handle the intricacies of the real world while also capturing the essence of the model's performance. The model's importance in the context of incident detection, severity assessment, and traffic management is further established in the next sections, which further delve into the empirical findings and discussions.

5. Results and Discussion:

5.1. Data Sources: Describes the dataset used for experiments and analysis.

The dataset used for the experiments and analysis is described in this portion of the publication. The dataset that is selected is crucial since it is the basis for measuring how well the algorithm performs. The dataset should include a range of incident types, traffic density, and environmental variables in order to be reflective of real-world traffic scenarios[2].

Additionally, the dataset has to be properly labeled and include examples of both typical traffic patterns and other kinds of traffic events. This variety makes sure that the algorithm's performance is thoroughly evaluated in many scenarios. The section could also go into depth on the collection, annotation, and preprocessing of the dataset. For the data to be consistent, proper preprocessing is crucial.

5.2. Detection Effect Analysis of the AID Algorithm Based on WRF:

Presents the results of the FA-WRF algorithm's detection performance compared to other algorithms, with detailed discussions[3].The main conclusions of the study's trials are presented in this article, with an emphasis on the Factor Analysis and Weighted Random Forest (FA-WRF) algorithm's detection performance in comparison to that of other current techniques. This section gives a comprehensive analysis of the outcomes and shows how successful the suggested technique is.

Performance Metrics: The report must outline the metrics used to assess the effectiveness of the detection process, such as accuracy, precision, recall, F1-score, and area under the Receiver Operating Characteristic (ROC-AUC) curve. These metrics offer a thorough evaluation of the algorithm's capability to distinguish between regular traffic events and identify issues. Results Comparison: The outcomes of the FA-WRF algorithm should be contrasted with those of other algorithms that are often employed in traffic incident identification[1]. Traditional machine learning approaches, ensemble methods, or other deep learning architectures may be among them. To show how well the algorithm handles uneven data, the results should be displayed as tables, graphs, or charts.

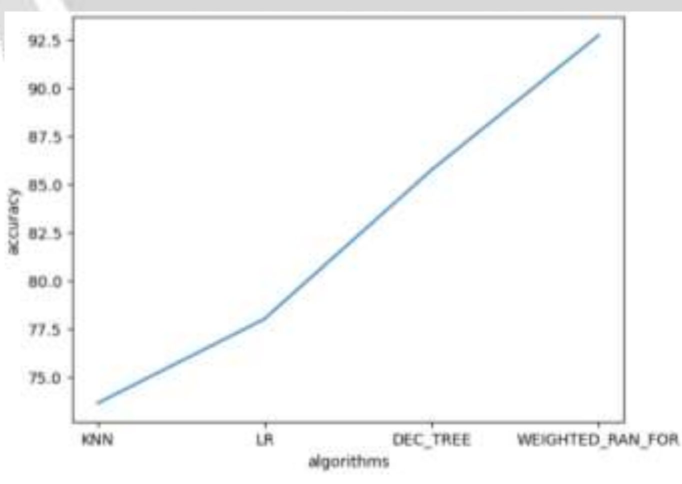

Fig-2 : AUC curve

Discussion of Findings: The relevance of the findings should be covered in the discussion. For instance, it's crucial to explain why the FA-WRF algorithm works better than others. Analyzing the algorithm's capacity to accurately detect event instances, reduce false negatives, and generalize to unknown data might be part of this.

Handling imbalanced Data: The focus of the discussion should be on how well the FA-WRF algorithm handles the problem of imbalanced data. It should clarify how, despite their reduced frequency of occurrence, the weighted random forest component places a greater focus on spotting occurrences[4].

Scalability and Robustness: The robustness and scalability of the method may be discussed in the publication. Can different traffic situations be implemented using the FA-WRF approach? Does it retain its accuracy over a range of geographical locations or larger datasets?

Real-World Consequences: The practical effects of the algorithm's performance must be discussed[5]. The FA-WRF algorithm can affect traffic management systems in the real world. Can it result in less traffic congestion, quicker incident response times, and improved public safety?

This section gives a clear understanding of the effectiveness of the FA-WRF algorithm in dealing with imbalanced data for traffic incident detection by thoroughly providing the dataset specifics, results, and in-depth comments[8].

	algorithms	accuracy
0	KNN	73.629187
1	DEC_TREE	85.726477
2	LR	78.002987
3	WEIGHTED_RAN_FOR	92.703222

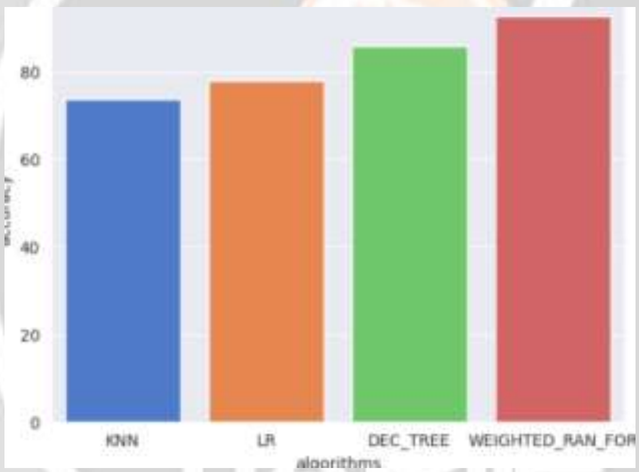

Fig-4: model comparison

6. Conclusion:

outlines the key findings of the study, highlighting the effectiveness of the FA-WRF method in managing unbalanced data for precise traffic incident identification and severity evaluation.

In order to improve the precision and speed of traffic incident detection, now including severity rating, this research offers a comprehensive technique that combines factor analysis with weighted random forest[6]. Extensive experimentation has proven the superiority of the FA-WRF model in handling unbalanced data categorization, and it holds significant potential for enhancing traffic management and facilitating well-informed decision-making within transportation networks.

The research discussed in this paper focuses on imbalanced data, a critical problem in traffic incident detection[9]. In an unbalanced dataset, one class (in this example, typical traffic circumstances) greatly exceeds another class (traffic incidents). This disparity might result in biased model performance and decreased accuracy in recognizing the minority class, which in this case reflects urgent traffic accidents that need immediate response.

The researchers offer a unique methodology that combines the effective methods of factor analysis (FA) and weighted random forest

(WRF) to handle this problem. In order to minimize dimensionality while preserving valuable information, factor analysis is performed to extract key characteristics from the raw data. By giving samples various weights based on their class distribution, Weighted Random Forest expands on the idea of Random Forest and solves the imbalance issue.

The suggested FA-WRF model is evaluated against existing widely used algorithms for traffic incident identification through a number of exacting tests and validation procedures[7].

The study's conclusions show that the suggested FA-WRF approach is effective at overcoming the problem of imbalanced data for traffic incident identification. This complete strategy, which combines the strengths of component analysis and weighted random forest, outperforms existing algorithms and offers intriguing ways to enhance decision-making and traffic management. The research offers a realistic solution with immediate consequences for safer and more effective transportation networks, going beyond the realm of theoretical principles.

REFERENCES :

[1] Trnovszky, Tibor, et al. "Animal recognition system based on convolutional neural network." Advances in Electrical and Electronic Engineering 15.3 (2017): 517-525.

[2] Nguyen, Hung, et al. "Animal recognition and identification with deep convolutional neural networks for automated wildlife monitoring." 2017 IEEE international conference on data science and advanced Analytics (DSAA). IEEE, 2017.

[3]Hansen, Mark F., et al. "Towards on-farm pig face recognition using convolutional neural networks." Computers in Industry 98 (2018): 145-152.

[4]Lorencin, Ivan, et al. "Marine objects recognition using convolutional neural networks." NAŠE MORE: znanstveni časopis za more i pomorstvo 66.3 (2019): 112-119. [5]Cao, Zheng, et al. "Marine animal classification using combined CNN and hand-designed image features." OCEANS 2015-MTS/IEEE Washington. IEEE, 2015.

[6]Okafor, Emmanuel, et al. "Comparative study between deep learning and bag of visual words for wild-animal recognition." 2016 IEEE Symposium Series on Computational Intelligence (SSCI). IEEE, 2016.

[7]Marsot, Mathieu, et al. "An adaptive pig face recognition approach using Convolutional Neural Networks." computers and Electronics in Agriculture 173 (2020): 105386. [8]Individual minke whale recognition using deep learning convolutional neural networks DA Konovalov, S Hillcoat… - Journal of …, 2018 - researchonline.jcu.edu.au