Volume: 16 Issue: 01 Pages: 6304– 6310 (2024) ISSN: 0975-0290

Enhancing Integrity of Toll Gates:Fastag Fraud Detection

Prathamesh Pradeep Dessai, Nikhil B, Dr. S. Senthil
Professor & DirectorSchool of CSA REVA University Bengaluru, India

ABSTRACT

The FastTag Fraud Detection System employs a machine learning model to identify fraudulent activities in FastTag transactions. Key features such as 'Transaction_Amount,' 'Amount_paid,' 'Vehicle_Type,' 'Lane_Type,' and'Geographical_Location' are used to differentiate betweenlegitimate and potentially fraudulent transactions. The modelconsiders various classifiers including Stochastic Gradient Descent (SGD), K-Nearest Neighbors (KNN), XGBoost, Logistic Regression, and Support Vector Machines (SVM). The SGD Classifier emerges as the most effective, demonstrating high accuracy, perfect precision, and abalanced recall-precision ratio. The model building process involves encoding categorical features, splitting the dataset, and training and evaluating multiple classifiers. Test accuracyfor models like Logistic Regression was 94.5%; test accuracyfor SGD Classifier was 98.0%; test accuracy for Gradient Boosting Classifier was 97.2%; test accuracy for SVC was 97.0%; and test accuracy for KNeighbors Classifier was 97.9%. These evaluation results demonstrated the efficacy ofthe FastTag Fraud Detection System. Strong precision, recall,and F1-score metrics were displayed by these models, demonstrating their capacity to precisely identify fraudulent transactions and improve the security of electronic toll collection systems.

Keywords: FastTag fraud detection, Toll gates, Fraudulent activities, electronic toll collection systems, Revenue protection

Date of Submission : May 08, 2024 Date of Acceptance : June 22, 2024

1. INTRODUCTION

A new age in toll gate operations has been brought about by the introduction of electronic toll collecting systems, typified by technologies like FastTag [1]. This has improved overall efficiency and streamlined the toll collection process. But these technological advances have also brought with them a new set of difficulties, especially when it comes to fraud that jeopardizes the integrity of toll gate revenue [2]. FastTag, a well-known electronic toll collecting system, has been the target of several fraudulent schemes in which people try to take advantage of gaps to avoid paying the exact toll amount. This makes it necessary to put in place a strong fraud detection system that is designed especially to protect the money made from toll collections and guarantee the dependability of the ecosystem supporting electronic toll payments [1].

Effective fraud detection procedures are becoming more and more important as electronic toll collecting systems continue to gain traction in the transportation sector [2]. Because of the vulnerabilities in the FastTag system, which include the possibility of mismatched vehicle types, server outages, RFID skimming, stolen tags, and collusion, it is crucial to take preventative action against fraudulent actions [1]. Toll collection authorities can maintain the integrity of toll gate income and reduce the risks associated with fraudulent transactions by addressing these technological constraints and human vulnerabilities. With the use of cutting-edge technology and analytical techniques, the proposed FastTag Fraud Detection System seeks to improve the security and effectiveness of toll collection procedures by successfully identifying and stopping fraudulent activity [15].

Accurately identifying differences in transaction data that can point to fraudulent activity is one of the main obstacles in the fight against FastTag fraud [1]. The suggested fraud detection system may analyze transaction patterns, identify abnormalities, and provide real-time alerts for suspected activity by utilizing advanced data analytics, machine learning algorithms, and artificial intelligence approaches. By taking a proactive stance, toll collection authorities can prevent financial losses, expedite the investigation of possible fraud cases, and maintain the credibility and confidence of the electronic toll collection system [15]. The FastTag Fraud Detection System provides a comprehensive solution to improve the security and dependability of toll gate operations by ongoing monitoring and analysis of transaction data, ultimately helping both FastTag users and toll collection authorities [1].

2. LITERATURE SURVEY

P. Arokianathan, V. Dinesh *et.al.* [1] proposes an automatic toll collection and theft detection system at toll plazas. The system uses RFID technology for vehicle identification and automatic toll collection, eliminating the need for manual toll collection. In addition, the system has a theft detection mechanism that uses sensors to detect attempts to tamper with toll plazas or steal tolls. This paper describes the system's architecture, hardware, and software components, as well as the algorithms used for toll collection and theft detection. The proposed system aims to improve the efficiency of toll collection, reduce congestion at toll plazas, and prevent toll theft.

Y. K. Al-Audah, A. K. Al-Juraifani et.al. [2] Achieve a success rate of 94% with a processing time of less than 40 Ms/plate under optimal conditions. The system uses morphological image processing and LabVIEW's built- in OCR tools for character recognition. Future improvements include fusion algorithms that combine Arabic and Latin characters, and statistical approaches to license plate recognition.

A. Y. Felix, A. Jesudoss et.al. [3] Addresses the challenge of accurately identifying license plates for vehicle tracking and entry/exit monitoring. The proposed system uses CCTV cameras, image processing tools, and various techniques such as adaptive histogram equalization (AHE), active contour method, optical character recognition (OCR), and deep neural network (DNN) for license plate recognition. and used for classification. Advantages of this system include low-cost digital cameras and improved methods for accurate license plate recognition. Suggested future improvements include the integration of cloud-based systems to improve data storage and retrieval, and the use of mobile agents for efficient data collection.

P. Patil, C. Kanagasabapathi et.al. [4] Designed for security applications. Processes the captured vehicle images using image identification techniques and his MATLAB, enabling efficient license plate extraction and recognition. The system meets the need for accurate, automated vehicle monitoring and offers possibilities for future improvements, including real-time processing, integration with cloud-based databases, and integration of machine learning algorithms to improve detection accuracy and speed. Offers.

Z. B. Musa and J. Watada [5] Presents a study on building a multi-camera tracking system for vehicle license plate recognition. The system consists of four main modules: image capture, preprocessing, vehicle location detection, and vehicle license plate location detection. The authors propose his two methods, a statistical method, and a transformation method, to detect the location of vehicle license plates. Statistical techniques include feature extraction, filtering, and location, while transformation techniques use clustering, filtering, and location based on histogram graphs. In this study, we compare the performance of these methods under different conditions such as normal images, rainy weather images, and low-quality images, and present experimental results demonstrating the effectiveness of each method. This paper provides an insight into the challenges and techniques in vehicle license plate recognition and provides a comparative analysis of the two proposed methods.

C. -H. Lin, Y. -S. Lin et.al. [6] Presents an efficient hierarchical license plate recognition system Overcome the challenges of traffic congestion and bad weather. Deep learning technologies such as YOLOv2 for vehicle recognition, SVM for license plate recognition, and LPRCNN for character recognition are used. to achieve high accuracy. Future improvements to the system may include advanced deep learning techniques to further improve detection accuracy and robustness in difficult situations.

Myung-Ryul Choi, Jin-Sung Park et.al. [7] As part of Intelligent Transportation Systems (ITS). AVIS consists of an invisible barcode, an optical scanner, and a DSP card. The optical scanner uses a 1350nm laser diode to read invisible barcode data, and a DSP board with TMS320C31 DSP chip processes the barcode data obtained from the optical scanner. The DSP card operates at 33 MHz and has error detection and correction capabilities. The system was successfully tested in the laboratory and on the test track, showing promising results. AVIS is expected to be applied to non-stop toll plazas and violation enforcement systems (VES) in the future. This article also describes field tests where the distance between the scanner and the barcode was 2.4 meters and the vehicle speed varied between 10 and 50 km/h. Test results show that AVIS recognizes approximately 98% of barcode data at 30 km/h. The detected data is sent to the host computer via an RS-232 line. The barcode attached to your car contains 15 digits of information such as license plate number, manufacturer, and color.

K. K. Kim, K. I. Kim et.al. [8] Presents about vehicle recognition, license plate segmentation using neural network filters, and license plate recognition using support vector machines (SVM). The system was evaluated on 1,000 video clips and achieved impressive performance with a 100% correct detection rate for vehicle detection, 97.2% for license plate segmentation, and 94.7% detection performance for license plate recognition. This study addresses the challenges of license plate recognition and suggests future improvements focused on improving system performance and integrating advanced deep learning techniques to further increase accuracy.

X. Tan, H. Wang et.al. [9] Presents a novel anti-collision algorithm for passive UHF RFID systems It uses signal recovery techniques to restore tag communications and obtain an accurate count of all tags in the field. This algorithm improves throughput by obtaining multiple valid communications from each collided slot with a DFSA-based anti-collision protocol, resulting in nearly 100% improvement. This paper also presents a systematic analysis of hardware-based anti-collision algorithms based on signal recovery and an evaluation of factors that influence system efficiency based on signal recovery. The authors present a practical hardware implementation of the proposed algorithm and provide a frame length optimization algorithm using a collision signal separation method. Potential applications for passive UHF RFID systems include supply chain management and electronic toll collection.

L. H. Godage and G. D. S. P. Wimalaratne [10] Has introduced an innovative application that identifies the front number plates of Sri Lankan vehicles using mobile devices. The system uses image manipulation, morphological manipulation, edge detection techniques, and character segmentation to achieve real- time recognition. By leveraging the OpenCV mobile

Volume: 16 Issue: 01 Pages: 6304– 6310 (2024) ISSN: 0975-0290

library and the Tesseract OCR library, the application processes image data and maintains a library of character templates in a mobile environment. This article describes the implementation of the system on a Sony Xperia m2 Aqua mobile device running the Android operating system, highlighting the use of the Eclipse Standard Android Software Development Kit for algorithm development and compilation. In experiments on 40 actual vehicles, we achieved a detection rate of 92.5%, considering issues such as dirt and roughness of panel fastening nails being mistaken for drawn edges. The flexibility of this system allows for instant generation of digital characters and tracking of geolocation, making it more adaptable than traditional CCTV-based detection systems. Overall, this paper presents a comprehensive approach to real-time mobile. license plate recognition that offers potential benefits for applications such as electronic toll collection, traffic monitoring, law enforcement, and urban parking access. is shown.

J. Watada and Z. B. Musa [11] Describes a multi-camera tracking system focused on detecting vehicle signs on highways. The system uses grid computing to overcometime-consuming processes and consists of multiple cameras,grid computing systems, databases, and networks. This paperalso provides an overview of related work in the field of multi-camera tracking systems and discusses future researchtrends.

S. Hemalatha and E. Prabhu [12] Describes the use of dedicated short-range communication (DSRC), FM0 coding, Manchester coding, and RFID technology in communication systems. The importance of DSRC in transportation communication and potential applications of DSRC in intelligent transportation systems are highlighted. This paper also describes a similarity-oriented logic simplification technique aimed at improving hardware utilization in DSRC systems. Additionally, this article considers the integration of RFID technology and intelligent transportation systems for electronic toll collection. He discusses the potential benefits of using RFID technology in various areas such as logistics, supply chain, race timing, and library systems. The structure of this paper includes sections on related research, coding principles, SOLS methodology, proposed Miller coding, results and discussion, conclusion, and future areas. The “Related Work” section provides insights into previous research on Manchester coding, secure vehicular communications, DSRC standards, and various applications of RFID technology.

A. Kumar, N. Anusha et.al. [13] Provides a comprehensive system that integrates various technologies to address transportation challenges. The proposed system aims to streamline toll collection, improve security through Breathalyzer detection, and monitor vehicle loading using IoT and postal systems. Key components of the system include RFID technology for automatic toll payment, an alcohol sensor (MQ135) to detect driver alcohol consumption, and a toll plaza load control sensor to monitor vehicle loading. RFID technology allows tolls to be automatically debited from a user's prepaid account when a vehicle passes through a toll plaza. The purpose of the alcohol sensor is to detect the amount of alcohol consumed by the driver and automatically brake the vehicle if the limit is exceeded. In addition, load control sensors ensure that heavy vehicles do not exceed allowable load limits, reducing penalties and automatically emailing commission personnel. System implementation includes the use of an MSP430 Launchpad microcontroller, RFID tags and readers, alcohol sensors, load control sensors, and an email system to send notifications and alerts to relevant parties. This paper also mentions relatedworks and literature that contribute to the understanding and development of the proposed system.

W. A. Syafei, M. A. Fatkhurrahman et.al. [14] Presents a prototype of a smart zero-queue toll plaza system based on wireless technology. The system is designed to quickly recognize vehicles and process payments in less than one second per vehicle, eliminating the need for paper receipts and reducing wait times in line. The main benefits of implementing this system are reduced traffic congestion and environmental impact. This paper also discusses existing toll plaza systems and their limitations, as well as the potential for future implementation of queuing-free smart toll plaza systems.

Y. Guo, H. Li et.al. [15] Describes how to apply RFID-SIM in electronic toll collection (ETC) systems to overcome the shortcomings of traditional ETC systems. We describe the system structure, hardware components, and the interaction process between the RFID-SIM and the RFID reader. The RFID SIM card embedded in the mobile phone acts as a vehicle unit, allowing automatic toll payment and access control. The study highlights the benefits of RFID SIMs, including convenience, reduced size, and improved security. It also highlights the use of electronic tags and conflict detection to manage transaction distance and prevent errors. The paper concludes that RFID SIM cards have great potential to be widely applied in ETC due to their convenience and intelligence.

3. METHODOLOGY

3.1 DATA SOURCES

Data collection: The process of Data collection included synthetic data creation, generating a dataset that simulates real world toll transactions. Also, as has been said, the data is synthetic data, with a total of 5000 records taken in the process and based on an Indian dataset from the state of Bengaluru. This dataset includes details such as Transaction ID, Timestamp, Vehicle Type, FASTag ID, Tollbooth ID, Lane Type, Vehicle Dimension, TransactionAmount, Amount Paid, Geographical Location, Vehicle speed,and Vehicle plate number. The aim is to create a realistic representation of both legitimate and fraudulent transactions inthe dataset, thereby simulating real-world scenarios. This comprehensive dataset serves as the foundation for developing and testing the robustness of the FASTag fraud detection system.

Volume: 16 Issue: 01 Pages: 6304– 6310 (2024) ISSN: 0975-0290

```
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   Transaction_ID        5000 non-null   int64
 1   Timestamp             5000 non-null   object
 2   Vehicle_Type          5000 non-null   object
 3   FastagID              4451 non-null   object
 4   TollBoothID           5000 non-null   object
 5   Lane_Type             5000 non-null   object
 6   Vehicle_Dimensions    5000 non-null   object
 7   Transaction_Amount    5000 non-null   int64
 8   Amount_paid           5000 non-null   int64
 9   Geographical_Location 5000 non-null   object
 10  Vehicle_Speed         5000 non-null   int64
 11  Vehicle_Plate_Number  5000 non-null   object
 12  Fraud_indicator       5000 non-null   object
dtypes: int64(4), object(9)
```

Fig 3.1: Figure representing the dataset components and their datatypes.

3.2 DATA PROFILING

The attributes that have been chosen cover a range of elements that are essential for identifying fraudulent conduct in toll booth transactions. The numerical indicators "Transaction_Amount" and "Amount_paid" are crucial since they can identify any disparities that can point to fraud. By providing insights into transaction settings, categorical variables such as 'Vehicle_Type' and 'Lane_Type' facilitate the identification of patterns linked to certain vehicle kinds or lane use. Furthermore, the geographic context provided by 'Geographical_Location' helps identify regional abnormalities or fraudulent activity patterns. Together, these carefully chosen characteristics provide a thorough understanding of transactions and make it easier to spot anomalies that might be signs of fraud. The model ensures optimal performance and efficiency by carefully selecting features that are crucial for successful fraud detection.

Feature Engineering: In the feature engineering step of building the model, a selection of key features is made that will be used for training. The features chosen for this model include 'Transaction_Amount', 'Amount_paid', 'Vehicle_Type', 'Lane_Type' and 'Geographical_Location'.

'Transaction_Amount' and 'Amount_paid' are numerical features that represent the amount of money involved in the transaction and the amount actually paid by the user, respectively. Discrepancies between these two values could potentially indicate fraudulent activity. 'Vehicle_Type' and 'Lane_Type' are categorical features that provide information about the type of vehicle involved in the transaction and the type of lane used at the toll booth. These features could help identify patterns or trends associated with certain types of vehicles or lanes.

Finally, 'Geographical_Location' is another categorical feature that indicates the location of the toll booth where the transaction took place. This feature could be useful in identifying geographical patterns in fraudulent activities.

These features were selected because they provide a comprehensive overview of the transaction and could potentially reveal patterns or anomalies indicative of fraud. The selection of appropriate features is a critical step in building an effective machine learning model.

3.3 DATA CLEANING AND PREPROCESSING

The preprocessing stage of the data involves several crucial steps. Firstly, any null values present in the FastagID column are replaced with 0. This is done to ensure that the dataset is free from missing values which could potentially interfere with the model's performance. Next, the timestamp is converted to a datetime format using the pandas library. This conversion is necessary to facilitate easier manipulation and analysis of the time-related data.

Finally, all categorical columns are converted into numerical columns using a label encoder. The columns that undergo this transformation include 'Vehicle_Type', 'FastagID', 'TollBoothID', 'Lane_Type', 'Vehicle_Dimensions', 'Geographical_Location', 'Vehicle_Plate_Number' and 'Fraud_indicator'. This step is essential as machine learning models generally work better with numerical data. These preprocessing steps are vital in preparing the data for the subsequent stages of model training and evaluation.

3.4 DATA MODEL

The development of a Classification Machine Learning model is crucial for detecting fraudulent activities in FASTag transactions. Leveraging key features such as 'Transaction_Amount,' 'Amount_paid,' 'Vehicle_Type,' 'Lane_Type,' and 'Geographical_Location,' the model aims to distinguish between legitimate transactions and those indicative of fraud.

a) Transaction Amount and Amount Paid: These features provide insight into the financial aspects of transactions. A classification model can discern unusual patterns in the relationship between the transaction amount and the amount paid, identifying potential discrepancies that may signal fraudulent activity.

b) Vehicle Type: 'Vehicle_Type' serves as a significant feature for classification. Certain vehicle types may be more susceptible to fraudulent activities, and the model can learn to recognize such patterns based on historical data.

c) Lane Type: 'Lane_Type' offers information about the specific type of lane used in a transaction. The model can discern whether certain types of lanes are more prone to fraudulent activities, contributing to the overall fraud detection strategy.

Volume: 16 Issue: 01 Pages: 6304– 6310 (2024) ISSN: 0975-0290

d) Geographical Location: 'Geographical_Location' provides context regarding where the transactions occur. The model can identify patterns of fraud associated with specific regions, enabling a more targeted detection approach.

3.5 MODEL SELECTION

a) SGD Classifier: The Stochastic Gradient Descent (SGD) Classifier is a linear classifier optimized by SGD. It's efficient and easy to implement, making it a good choice for large-scale and sparse machine learning problems. In the context of FastTag fraud detection, it can be used to draw linear boundaries to distinguish between fraudulent and non-fraudulent transactions.

b) KNN Classifier: The K-Nearest Neighbors (KNN) Classifier is a type of instance-based learning that classifies new instances based on their similarity to existing instances in the training set. It can be useful in detecting FastTag fraud by identifying transactions that are like known fraudulent transactions.

c) Gradient Boosting: Gradient Boost stands for Gradient Boosting. It's an implementation of gradient boosting machines that's known for its speed and performance. Gradient Boost can handle a variety of data types, missing values, and is great at avoiding overfitting. It can be used to identify complex patterns that may indicate FASTag fraud.

d) Logistic Regression: Logistic Regression is a statistical model used for binary classification problems. It estimates the probability of an event occurring. In the case of FASTag fraud detection, it can estimate the probability of a transaction being fraudulent based on the given features.

e) SVM: Support Vector Machines (SVM) are powerful models used for both classification and regression. They work by finding a hyperplane that best separates the classes. SVMsare effective in high-dimensional spaces, making them usefulfor FASTag fraud detection if you have many features.

4. RESULTS

Table 4.1 Presents the performance of several machine learning models on a classification task. The models are listed in the first column, labeled "Model". The second and third columns are labeled "Train Accuracy" and "Test Accuracy", respectively. Train accuracy is the proportion of times the model correctly classified examples from the training data. Test accuracy is the proportion of times the model correctly classified examples from a separate test data set. In an ideal scenario, the train and test accuracy should be very similar. A large difference between train and test accuracy suggests that the model may be overfitting the training data.

Five distinct machine learning methods were trained and assessed on a dataset for this classification. Training accuracy of 95.625% and testing accuracy of 94.5% were attained by the logistic regression model. Because of its simplicity and interpretability, logistic regression is a linear classifier that is frequently used to estimate the likelihood of a binary result. Stochastic gradient descent optimization was used by the SGD classifier, which performed even better, with testing accuracy of 98.0% and training accuracy of 98.3%. With a training accuracy of 96.775% and a testing accuracy of 97.2%, the powerful ensemble learning technique known as the gradient boosting classifier demonstrated its capacity to systematically improve model predictions by combining weak learners. The support vector classifier (SVC) used the idea of locating the ideal hyperplane to divide classes in feature space to achieve training accuracy of 97.525% and testing accuracy of 97.0%. Finally, the k-nearest neighbors KNeighbors classifier, which bases predictions on the majority class among the closest neighbors in feature space, had the greatest testing accuracy of 97.9% and the highest training accuracy of 99.25%. Based on variables like accuracy, interpretability, and computing efficiency, these data demonstrate the heterogeneous performance of different classifiers and provide insight into which ones are best suited for a certain classification assignment.

Table 4.1: Comparison of Machine Learning Model Performance

Model	**Train Accuracy**	**Test Accuracy**
Logistic Regression	95.625	94.5
SGD Classifier	98.300	98.0
Gradient Boosting Classifier	96.775	97.2
SVC	97.525	97.0
KNeighbors Classifier	99.250	97.9

Table 4.2 Presents all five machine learning models—logistic regression, SGD classifier, Gradient Boosting classifier, SVC, and KNeighbors classifier—showed remarkable accuracy in this classification assignment, each getting a flawless 100% score. This exceptional accuracy shows that the models can correctly categorize instances into the appropriate classes. While the SGD Classifier performed exceptionally well in terms of efficiency and scalability, logistic regression demonstrated strong performance despite its simplicity. The effectiveness of ensemble learning was exhibited by the Gradient Boosting Classifier, but the SVC was found to be efficient in high-dimensional areas. Like this, the KNeighbors Classifier produced accurate classification since it relied on nearest neighbors. All things considered, these models provide trustworthy answers for a range of classification tasks, guaranteeing precise predictions with high degrees of confidence.

Table 4.2: Precision Scores of Different Classifiers.

Model	**Precision**
Logistic Regression	100
SGD Classifier	100

Gradient Boosting Classifier	100
SVC	100
KNeighbors Classifier	100

Table 4.3 the recall ratings for the following classifiers are shown in the table: KNeighbors Classifier, SGD Classifier, Gradient Boosting Classifier, SVC, and Logistic Regression. Recall, often referred to as sensitivity, represents the percentage of real positive examples that the classifier correctly detected. Recall scores were greatest for the SGD Classifier 90.78%, and KNeighbors Classifier 90.32% were not far behind. Both the SVC and the Gradient Boosting Classifier showed excellent recall rates, scoring 86.18% and 87.10%, respectively. At 74.65%, the Logistic Regression model's recall score was much lower. Higher recall values indicate greater success in identifying true positives. These recall values provide insights into the classifiers' capacity to recognize positive cases.

Table 4.3: Recall Scores of Different Classifiers.

Model	Recall
Logistic Regression	74.654378
SGD Classifier	90.783410
Gradient Boosting Classifier	87.096774
SVC	86.175115
KNeighbors Classifier	90.322581

Table 4.4 provides details about the F1 scores of several machine learning models, such as KNeighbors Classifier, Gradient Boosting Classifier, SGD Classifier, and Logistic Regression. The F1 score is a statistic that provides a fair assessment of a model's performance by combining recall and accuracy. Better overall performance in terms of memory and accuracy is indicated by higher F1 scores. With an F1 score of 95.17%, the SGD Classifier outperformed the Gradient Boosting Classifier, which came in second place with a score of 93.10%. Based on these findings, it appears that the SGD Classifier and Gradient Boosting Classifier are better at balancing precision and recall in their predictions, which may make them good options for jobs where it's important to strike a balance between false positives and false negatives.

Table 4.4: Table represents the F1 score.

Model	F1 score
Logistic Regression	85.488127
SGD Classifier	95.169082
Gradient Boosting Classifier	93.103448
SVC	92.574257
KNeighbors Classifier	90.322581

5. CONCLUSION AND SCOPE FOR FUTURE ENHANCEMENT

In conclusion, the developed Classification Machine Learning model for the FASTag Fraud Detection System effectively distinguishes between legitimate and fraudulent transactions. Using key features like 'Transaction_Amount,' 'Amount_paid,' 'Vehicle_Type,' 'Lane_Type,' and 'Geographical_Location,' the model underwent a rigorous selection process, with the SGD Classifier emerging as the top performer among contenders like KNN, Gradient Boosting, Logistic Regression, and SVM. The model-building phase included thorough preprocessing, feature selection, and evaluation, with the KNeighbors Classifier demonstrating the highest training and test accuracy. However, the SGD Classifier excelled with exceptional overall performance, achieving high accuracy, precision, recall, and F1-score. This approach positions the SGD Classifier as the most suitable choice for the FASTag Fraud Detection System, providing reliability in identifying fraudulent activities and serving as a robust framework applicable to enhancing the security of electronic toll collection systems, contributing to the ongoing fight against FastTag transaction fraud.

Looking ahead, there are several key areas for future enhancement. Firstly, advancing threat intelligence and pattern recognition by integrating cutting-edge AI and machinelearning techniques could enhance analytical capabilities. This could involve leveraging graph networks to model complex transaction relationships or employing unsupervised learning algorithms to uncover hidden patterns indicative of new fraud schemes. Additionally, integrating frictionless biometric authentication mechanisms like iris or voice recognition could bolster user verification and transaction security without impacting user experience. Exploring federated learning approaches could enable collaborative fraud detection across toll operators without compromising sensitive user data privacy. These future directions could significantly enhance the efficacy of the FASTag Fraud Detection System and further improve electronic toll collection system security.

REFERENCES

[1] P. Arokianathan, V. Dinesh, B. Elamaran, M. Veluchamy and S. Sivakumar, ***"Automated toll booth and theft detection system"*** 2017 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR), Chennai, India, 2017, pp. 84-88, doi: 10.1109/TIAR.2017.8273691.

[2] Y. K. Al-Audah, A. K. Al-Juraifani and M. A. Deriche ***"A real-time license plate recognition system for Saudi Arabia using LabVIEW"*** 2012 3rd International Conference on Image Processing Theory, Tools and Applications (IPTA), Istanbul, Turkey, 2012, pp. 160-164, doi: 10.1109/IPTA.2012.6469576.

[3] A. Y. Felix, A. Jesudoss and J. A. Mayan, ***"Entry and exit monitoring using license plate recognition"*** 2017 IEEE International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), Chennai, India, 2017, pp. 227-231, doi: 10.1109/ICSTM.2017.8089156.

[4] P. Patil, C. Kanagasabapathi and S. S. Yellampalli ***"Automatic number plate recognition system for vehicle***

identification" 2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT), Mysuru, India, 2017, pp. 431-434, doi: 10.1109/ICEECCOT.2017.8284543.

[5] Z. B. Musa and J. Watada, ***"Statistical Versus Transform Methods: An Experimental Comparison of Vehicle Plate Location Detection for a Multi-Camera Tracking System"*** Second International Conference on Innovative Computing, Information and Control (ICICIC 2007), Kumamoto, Japan, 2007, pp. 421-421, doi: 10.1109/ICICIC.2007.540.

[6] C. -H. Lin, Y. -S. Lin and W. -C. Liu, ***"An efficient license plate recognition system using convolution neural networks"*** 2018 IEEE International Conference on Applied System Invention (ICASI), Chiba, Japan, 2018, pp. 224-227, doi: 10.1109/ICASI.2018.8394573.

[7] Myung-Ryul Choi, Jin-Sung Park, Seoul-Sun Lee and Seung-Ho Tak, ***"Automatic vehicle-identification system"*** Proceedings of Digital Processing Applications (TENCON '96), Perth, WA, Australia, 1996, pp. 68-72 vol.1, doi: 10.1109/TENCON.1996.608709.

[8] K. K. Kim, K. I. Kim, J. B. Kim and H. J. Kim, ***"Learning-based approach for license plate recognition"*** Neural Networks for Signal Processing X. Proceedings of the 2000 IEEE Signal Processing Society Workshop (Cat. No.00TH8501), Sydney, NSW, Australia, 2000, pp. 614-623 vol.2, doi: 10.1109/NNSP.2000.890140.

[9] X. Tan, H. Wang, L. Fu, J. Wang, H. Min and D. W. Engels, ***"Collision Detection and Signal Recovery for UHF RFID Systems"*** in IEEE Transactions on Automation Science and Engineering, vol. 15, no. 1, pp. 239-250, Jan. 2018, doi: 10.1109/TASE.2016.2614134.

[10] L. H. Godage and G. D. S. P. Wimalaratne, ***"Real-Time Mobile Vehicle License Plates Recognition in Sri Lankan Conditions"*** 2019 14th Conference on Industrial and Information Systems (ICIIS), Kandy, Sri Lanka, 2019, pp. 146-150, doi: 10.1109/ICIIS47346.2019.9063258.

[11] J. Watada and Z. B. Musa, ***"A Future View of a Multi-Camera Tracking System"*** in "Organized Session: SICE City" 2006 SICE-ICASE International Joint Conference, Busan, Korea (South), 2006, pp. 4870-4876, doi: 10.1109/SICE.2006.314610.

[12] S. Hemalatha and E. Prabhu, ***"An Analysis of Cognitive Radio Spectrum Sensing Methods in Communication Networks"*** 2018 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 2018, pp. 0462-0467, doi: 10.1109/ICCSP.2018.8524444.

[13] A. Kumar, N. Anusha and B. S. S. V. Prasad, ***"Automatic toll payment, alcohol detection, load and vehicle information using Internet of things & mailing system"*** 2017 International Conference on Intelligent Computing and Control (I2C2), Coimbatore, India, 2017, pp. 1-5, doi: 10.1109/I2C2.2017.8321775.

[14] W. A. Syafei, M. A. Fatkhurrahman, Y. Baharuddin, A. Hidayatno, Darjat and H. Ochi, ***"Queuing Free Smart Toll Gate based on Wireless Technology"*** 2018 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), Ishigaki, Japan, 2018, pp. 24-30, doi: 10.1109/ISPACS.2018.8923140.

[15] Y. Guo, H. Li and Z. Li, ***"The Application of RFID-SIM in Electronic Toll Collection System"*** 2011 International Conference on Computational and Information Sciences, Chengdu, China, 2011, pp. 913-916, doi: 10.1109/ICCIS.2011.275.

[16] N. N. Thilakarathne et al., “Internet of Things (IoT) Security: Status, Challenges and Countermeasures,” International Journal of Advanced Networking and Applications, vol. 14, no. 03, pp. 5444–5454, Jan. 2022, doi: 10.35444/ijana.2022.14305.

[17] O. E. Olasupo, “An Efficient Vehicle-to-Everything (V2X) Communication Algorithm for the Deployment and Operation of Self-driving Cars,” International Journal of Advanced Networking and Applications, vol. 14, no. 01, pp. 5237–5249, Jan. 2022, doi: 10.35444/ijana.2022.14101.