

Identifying Factors Affecting Errors in Automatic Number Plate Recognition (ANPR) Systems Using the Random Forest Model: A Case Study of Urban ANPR Cameras in Mashhad

Mohammad Amin Ebrahimzadeh^{1*}, Mahdiseh Dabbaqi², Navid Nakhaei²

¹- Department of Civil Engineering, Faculty of Construction Engineering and Industrial Technologies, Islamic Azad University, Karaj Branch, Iran.

²- Department of Civil Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran.

Abstract

Automatic Number Plate Recognition (ANPR) cameras are among the key components of Intelligent Transportation Systems (ITS), playing a vital role in traffic monitoring, traffic violation enforcement, and the generation of traffic data for urban transportation management. However, recognition errors may reduce the quality and reliability of traffic data, leading to inaccurate planning and traffic management decisions. This study aims to identify and analyze the factors affecting errors in urban ANPR systems and to evaluate the capability of machine learning models in predicting these errors. The dataset was collected from two selected ANPR cameras in Mashhad, Iran, during the first week of June 2020. The input variables included camera ID, day of the week, recording time, vehicle speed, traffic lane, and traffic volume, while the dependent variables consisted of three types of errors: license plate recognition error, license plate character recognition error, and odd–even license plate classification error. To address the class imbalance problem, the Synthetic Minority Over-sampling Technique (SMOTE) was employed. The performance of the Random Forest (RF) and Artificial Neural Network (ANN) models was evaluated using separate training and testing datasets. The results revealed that recording time was the most influential factor affecting both license plate recognition errors and license plate character recognition errors. The Random Forest model outperformed the Artificial Neural Network, achieving classification accuracies of 85% and 86%, respectively. For odd–even license plate classification errors, traffic lane was identified as the most important predictor, and the Random Forest model achieved the highest accuracy of 96.5%. The findings demonstrate that the Random Forest algorithm is an effective tool for identifying error patterns and analyzing the factors influencing the performance of ANPR systems. The proposed approach can contribute to improving traffic data quality and enhancing the reliability of Intelligent Transportation Systems. Nevertheless, since the analysis was limited to two ANPR cameras and a one-week observation period, further studies using larger datasets collected from different locations and longer time periods are recommended to improve the generalizability of the findings.

Keywords: Automatic Number Plate Recognition (ANPR); Random Forest; Intelligent Transportation Systems (ITS); License Plate Recognition Errors; Urban Traffic Management

* Corresponding Author: Mohammad Amin Ebrahimzadeh:m.ebrahimzadeh@iau.ir

1. Introduction

Rapid urbanization, increasing vehicle ownership, and the growing complexity of travel patterns have made intelligent traffic management an essential component of modern urban planning. Intelligent Transportation Systems (ITS) have become indispensable tools for improving traffic operations, enhancing road safety, reducing congestion, and supporting evidence-based transportation planning. Among the various ITS technologies, Automatic Number Plate Recognition (ANPR) systems play a critical role in traffic surveillance, law enforcement, parking management, toll collection, and traffic data acquisition.

Accordingly, this study aims to identify and rank the factors influencing three major categories of ANPR errors, namely license plate recognition errors, license character recognition errors, and odd-even plate classification errors. Using real operational data collected from urban ANPR cameras in Mashhad, Iran, the predictive performance of the Random Forest algorithm is compared with that of an Artificial Neural Network (ANN). The findings provide valuable insights for improving ANPR reliability and enhancing the quality of traffic information used in Intelligent Transportation Systems.

2. Methodology

This research adopts a data-driven classification approach to investigate the factors affecting ANPR system errors. The dataset was obtained from two operational ANPR cameras installed along Shahid Kalantari Highway in Mashhad during the first week of June 2020.

Six explanatory variables were considered, including camera identification number, day of the week, recording time, vehicle speed, traffic lane, and traffic volume. Three binary dependent variables were defined representing the occurrence or absence of (1) license plate recognition error, (2) license character recognition error, and (3) odd-even plate classification error.

Because the dataset exhibited severe class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) was employed to balance the minority and majority classes before model development. Subsequently, the dataset was divided into training (70%) and testing (30%) subsets using the Hold-Out validation method.

Random Forest was selected as the primary classification algorithm because of its capability to model nonlinear relationships, reduce overfitting, and estimate variable importance. The model consisted of 500 decision trees using the Gini impurity criterion for node splitting. The number of variables considered at each split was determined as the square root of the total number of predictors, while the random seed was fixed at 42 to ensure reproducibility.

For comparison purposes, a feed-forward Artificial Neural Network with one hidden layer was also

developed. Model performance was evaluated using classification accuracy, confusion matrices, Receiver Operating Characteristic (ROC) curves, Area Under the Curve (AUC), and Mean Decrease Gini values for variable importance analysis.

3. Results

The Random Forest algorithm demonstrated superior predictive performance across all three classification problems compared with the Artificial Neural Network.

For license plate recognition errors, the Random Forest model achieved an overall classification accuracy of 85%. Variable importance analysis revealed that recording time was the most influential predictor, followed by traffic volume, vehicle speed, and camera identification. The ROC analysis yielded an AUC of approximately 0.80, indicating satisfactory discrimination between error and non-error observations.

For license character recognition errors, the Random Forest classifier achieved an accuracy of 86%. Similar to the previous model, recording time remained the dominant predictor. The close agreement between the training and testing AUC values demonstrated the model's robustness and generalization capability.

The highest predictive performance was obtained for odd-even plate classification errors, where the Random Forest model reached an accuracy of 96.5% with an AUC exceeding 0.95. Unlike the other error categories, traffic lane emerged as the most significant predictor, indicating that camera viewing geometry and vehicle positioning within different lanes substantially affect recognition performance.

Overall, the Random Forest algorithm consistently outperformed the Artificial Neural Network in terms of classification accuracy, stability, and predictive reliability, confirming its suitability for analyzing complex nonlinear relationships in operational ANPR datasets.

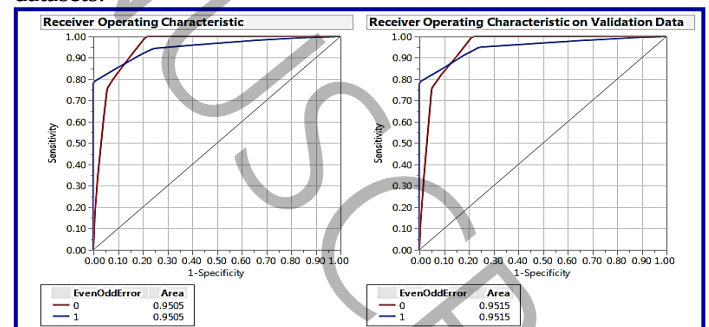


Figure 1. ROC curves for training and testing data in the random forest model

4. Discussion

The findings demonstrate that operational and traffic-related variables significantly influence the performance of ANPR systems under real-world conditions. Among all investigated variables, recording time exerted the greatest influence on both license plate

and license character recognition errors. This observation can be explained by changes in lighting conditions throughout the day, including direct sunlight, reflections from vehicle surfaces, reduced nighttime illumination, and fluctuations in image quality.

Traffic volume and vehicle speed also contributed substantially to recognition errors. Higher traffic demand increases vehicle interactions, image occlusion, and motion blur, all of which reduce the accuracy of optical character recognition algorithms.

For odd-even plate classification errors, traffic lane was identified as the dominant factor. This result suggests that lane-specific characteristics, including viewing angle, camera alignment, vehicle trajectory, and distance from the camera, strongly affect image quality and consequently the accuracy of license plate recognition.

The results are consistent with previous ITS research indicating that ensemble learning algorithms generally outperform individual machine-learning models when dealing with nonlinear and heterogeneous transportation datasets. The findings further indicate that improving ANPR performance requires not only advanced recognition algorithms but also optimized camera placement, appropriate installation geometry, regular equipment maintenance, and improved illumination conditions.

5. Conclusion

This study investigated the factors affecting Automatic Number Plate Recognition (ANPR) errors using real traffic data collected from urban cameras in Mashhad and compared the predictive capabilities of Random Forest and Artificial Neural Network models.

The results demonstrate that the Random Forest algorithm consistently provides superior predictive performance for all three categories of ANPR errors. Recording time was identified as the most influential variable affecting license plate and license character recognition errors, whereas traffic lane had the strongest impact on odd-even plate classification errors.

These findings suggest that operational conditions have a substantial influence on ANPR performance and should be considered alongside algorithmic improvements. The developed Random Forest model can therefore serve as an effective decision-support tool for monitoring ANPR performance, identifying error patterns, and improving the reliability of traffic databases used in Intelligent Transportation Systems.

One limitation of this study is that the analysis was based on data collected from only two cameras during a one-week observation period. Consequently, caution should be exercised when generalizing the findings to other locations or environmental conditions. Future research should incorporate longer observation periods, larger camera networks, additional environmental variables, and image-quality indicators to develop more comprehensive predictive models capable of further reducing ANPR recognition errors.

6. References

1. KhQadri MT, Asif M. Automatic number plate recognition system for vehicle identification using optical character recognition. *International Conference on Education Technology and Computer*, 17, pp. 335-338 (2009).
2. Roy A, Ghoshal DP. Number Plate Recognition for use in different countries using an improved segmentation. *National Conference on Emerging Trends and Applications in Computer Science*, pp1-5 (2011).
3. Tan H, Feng G, Feng J, Wang W, Zhang YJ, Li F. A tensor-based method for missing traffic data completion. *Transportation Research Part C: Emerging Technologies*, 1;28:15-27(2013).
4. Gaber T, Abdelwahab S, Elhoseny M, Hassanien AE. Trust-based secure clustering in WSN-based intelligent transportation systems. *Computer Networks*, 9;146:151-8(2018).
5. Robinson A, Venter C. Validating traffic models using large-scale Automatic Number Plate Recognition (ANPR) data. *Journal of the South African Institution of Civil Engineering*, 61(3):45-57.0(2019).
6. Gunay B, Saltan M, Ilgaz A, Albayrak Y. Area-wide ANPR coverage with a small number of instrumented vehicles. *International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, pp. 379-386(2015).
7. Homayounfar, Abdollah, Anthony TS Ho, Naida Zhu, George Head, and Phil Palmer. Multi-vehicle convoy analysis based on ANPR data (2011).
8. Anand SS, Büchner AG. Decision support using data mining (1998).
9. Soole D, Fleiter J, Watson B. Point-to-point speed enforcement [Austroads Project No.SS1649]. Austroads Ltd (2012).
10. Patel, C., Shah, D., Patel, A. Automatic number plate recognition system (ANPR): A survey. *International Journal of Computer Applications*, 69(9) (2013).
11. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P. The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM*, 39(11), 27-34 (1996).
12. Rigatti, S. J. Random forest. *Journal of Insurance Medicine*, 47(1), 31-39 (2017).