Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 14, No. 3, 2025

Vibration Feature Fusion and Machine Learning Based Road Surface Identification Method

Yuyan Zhu *

International Education College, Zhengzhou University of Light Industry, Zhengzhou Henan, 450002, China

*** Corresponding author Email:** 13938590396@163.com

Abstract: To address the demand for real-time road condition monitoring in intelligent transportation systems, a pavement roughness classification method based on multi-feature fusion and machine learning is proposed. By constructing vibration datasets for four typical types of road surfaces, 17-dimensional feature vectors including time-domain, frequency-domain, and statistical features are extracted. Feature importance evaluation and classification are performed using Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and other algorithms. Experimental results show that the proposed method achieves a classification accuracy of over 95.2% on the test set. Random Forest outperforms comparative methods in both training efficiency and feature interpretability, providing reliable technical support for real-time road condition monitoring and automated classification.

Keywords: Road Unevenness; Vibration Signal Analysis; Feature Extraction; Machine Learning; Real-time Classification.

1. Introduction

Road surface unevenness is a core indicator for assessing road quality and safety, directly affecting driving smoothness, ride comfort, and traffic safety [1-2]. Developing efficient and accurate detection and classification techniques is of great significance for promoting the transformation of road maintenance toward scientific and intelligent approaches.

Traditional methods, such as laser profilers, offer high precision but are limited by expensive equipment, low efficiency, and difficulties in meeting the demands of large-scale routine inspections, often requiring traffic closure[3]. In contrast, indirect detection methods based on vehicle vibration signals—which infer road surface conditions by collecting signals such as acceleration—have become a current research focus due to their low cost, high efficiency, and minimal disruption to traffic[4-5].

Previous research in technological development primarily utilized algorithms such as support vector machines[6-7] and random forests[8-10], combined with time-domain and frequency-domain analyses, to extract manually designed features such as root mean square (RMS) values and kurtosis for classification. However, the design of these features relies on expert experience and struggles to comprehensively capture the nonlinear patterns present in the signals. On the other hand, deep learning models such as Convolutional Neural Networks (CNN) [11-12] and Transformers [13-14], which have been introduced in recent years, are capable of automatically learning high-level features from raw signals, significantly enhancing recognition performance and generalization capabilities.

This field faces challenges such as high labeling costs, scarce real-world data, and the lack of systematic platforms. To address this, this paper proposes a classification system based on a parameterized road surface model and machine learning. The system generates vibration data at low cost, employs Random Forest for feature selection and classification, and builds an interactive interface that integrates simulation, prediction, and visualization, offering an integrated solution for intelligent maintenance.

2. System Architecture Design

A complete pavement classification workflow is designed, encompassing data generation, feature extraction, model training, and real-time prediction. The system first generates simulated vibration signals based on pavement types, capturing their physical distinctions. It then extracts features from the time domain, frequency domain, and statistical characteristics. These features are used for classification via a trained machine learning model, enabling real-time feature analysis and intelligent adjustment. This approach enhances the accuracy and reliability of online predictions.

2.1. Vibration Signal Model

The pavement vibration signal is composed of superimposed fundamental waves, high-frequency noise, and shock waves. The vibration parameters for the four road conditions (smooth, slightly uneven, uneven, severely uneven) exhibit regular variations: the fundamental frequency gradually increases from 1.0~15 Hz to 12.0~150 Hz; the amplitude rises from 0.02~0.05 to 0.7~1.0; the noise level increases from 0.005~0.01 to 0.15~0.25; the shock intensity grows from 0 to 0.5~0.7; the shock count increases from 0 to 6~9 times; and the high-frequency proportion rises from 0.01 to 0.30.

These parameter variations indicate that as road conditions deteriorate, the time required for vibration energy to dissipate increases, while the stimulation frequency rises and the attenuation slows down. This parametric modeling provides a clear basis for extracting core features (such as peak count and shock index), transforming the physical characteristics of the pavement into distinguishable mathematical features. This establishes a reliable data foundation for subsequent high-precision classification.

2.2. Feature Extraction and Selection

2.2.1. Multidimensional Feature Extraction

The system extracts and comprehensively characterizes road conditions through multidimensional features. In the time domain, features including root mean square, standard deviation, peak value, mean absolute deviation, and skewness are extracted to reflect energy and impact characteristics. In the frequency domain, spectral centroid, spread, and kurtosis are obtained to describe the frequency distribution via Fast Fourier Transform (FFT). Simultaneously, statistics such as peak count, average peak height, and their standard deviation are analyzed to quantify impact events, and the energy proportions in low, medium, and high-frequency bands are calculated.

2.2.2. Feature Importance Analysis

The importance of each feature is evaluated using the Random Forest algorithm. The ranking results are as shown. The top eight most important key features in descending order are: Feature importance analysis revealed that RMS was the most influential feature with an importance score of 0.145, followed by peak count (0.132) and spectral centroid (0.121). The remaining features, in descending order of importance, were: standard deviation (0.098), shock index (0.087), energy in the 8~15 Hz frequency band (0.076), crest factor (0.065), and peak-to-RMS ratio (0.054).

2.3. Model Training

2.3.1. Dataset Partitioning and Evaluation Metrics

(1) The dataset covers a total of 80,000 samples across four types of road surfaces, with 20,000 samples per type. It is split into training and test sets in a 7:3 ratio, with a fixed random seed to ensure balanced distribution and reproducible results, thereby avoiding class bias.

(2) The evaluation framework includes accuracy, precision, recall, and the F1-score, which respectively reflect overall classification performance, prediction reliability, sample recognition coverage, and the comprehensive performance of the model.

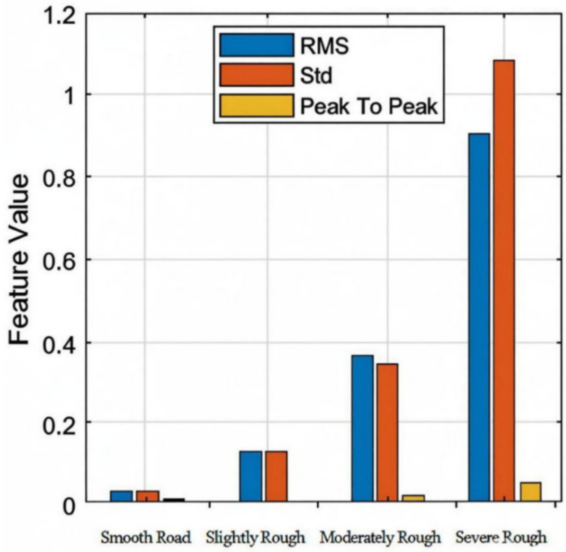
2.3.2. Comparative Analysis of Classification Algorithms

This study compares three classifiers: Random Forest, SVM, and KNN. All three achieved a weighted F1-score of 95.2%, validating the effectiveness of the extracted features. During training, Random Forest required 8.23 seconds and performed prediction in 1.56 milliseconds per sample; SVM training took 12.45 seconds with a prediction time of 0.89 milliseconds per sample; KNN had the fastest training time (0.35 seconds) but the slowest prediction speed (2.34 milliseconds per sample). Considering both prediction speed and model interpretability, Random Forest was ultimately selected.

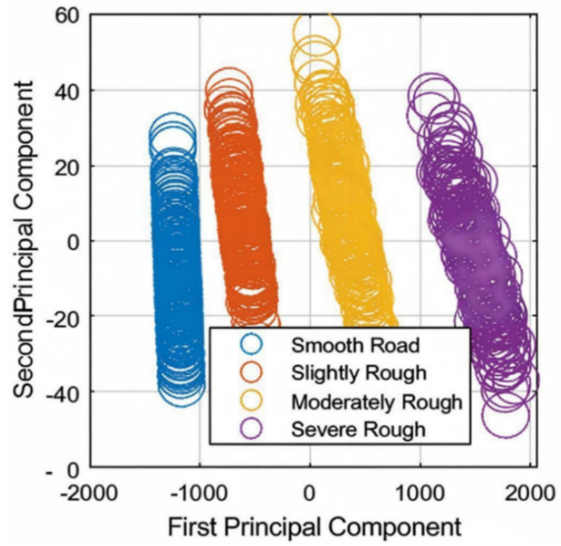
2.3.3. Visualization and Cluster Analysis

Visualization and cluster analysis are key for understanding the data and evaluating the model. Model performance and data characteristics are illustrated through the visualization of feature importance maps, projection visualizations via Principal Component Analysis (PCA), and visualizations of RMS value distributions.

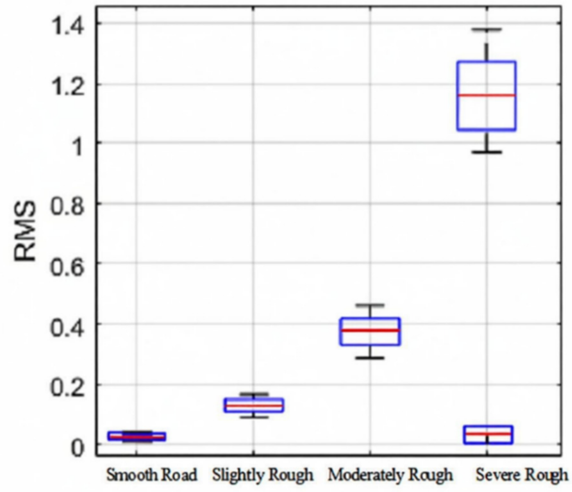
Visualization analysis indicates that the mean values of key features exhibit a stepwise increase as road surface unevenness intensifies, as shown in Figure 1(a). PCA dimensionality reduction reveals that samples from the four types of road surfaces form distinct and separate clusters, illustrated in Figure 1(b). Furthermore, the distributions of RMS values do not overlap and display a monotonic increase across categories, depicted in Figure 1(c). These findings confirm that the proposed features can effectively distinguish between different grades of road surfaces.



(a)Comparison of Mean Key Features



(b)PCA visualization shows perfect clustering



(c)RMS Distribution

Fig 1. Pavement Data Exploration and Differentiation

2.3.4. Analysis of Feature Differences Between Classes

The patterns and degree of differences between categories can be effectively determined through feature difference analysis. The analysis indicates that key feature values increase in a stepwise manner with rising category grades, with feature values of higher-grade groups being more than ten times greater than those of lower-grade groups (Fig.1(a)). In PCA visualization, different categories form distinct, nonoverlapping clusters in the principal component space, demonstrating significant separability (Fig.1(b)). The box plots of RMS distributions show no overlap between groups, and the small within-group variance indicates strong inter-class separability of the features (Fig.1(c)).

3. Real-time Road Surface Smoothness Prediction and Visualization Interface

A real-time road surface smoothness prediction system has been developed based on MATLAB, integrating signal analysis, machine learning, and visualization capabilities. The system monitors vehicle vibration signals and includes modules for data management, model training, and real-time demonstration, enabling the real-time identification of four types of road surfaces. Innovatively, the system incorporates a millisecond-level real-time prediction mechanism based on feature thresholds (e.g., RMS, spectral centroid) and an intelligent correction mechanism, significantly enhancing robustness. With dynamic visualization and status feedback, the interface adopts an industrial-grade design to ensure intuitive interaction.

4. Conclusion

(1) A MATLAB-based road surface roughness classification system has been developed. It generates vibration data for four types of road surfaces using a parametric model, identifies eight key features with class-wise hierarchical differences, and performs classification analysis on the data.

(2) Among 80,000 samples, multiple machine learning models achieved a classification accuracy of 95.2%. Through comprehensive comparison, the Random Forest algorithm not only ensures high classification performance but also maintains good model interpretability.

Acknowledgments

This work was supported by the Zhengzhou University of Light Industry Undergraduate Science and Technology Activity Project (2025-YB-0045) and the Henan Province College Student Innovation Training Program Project (202510462044).

References

[1] Wu D .Ride Comfort Prediction on Urban Road Using Discrete Pavement Roughness Index[J].Applied Sciences, 2024, 14(7): 3108-.DOI:10.3390/APP14073108.

[2] Patricia A P ,Henri G .A recent overview of the effect of road surface properties on road safety, environment, and how to monitor them.[J].Environmental science and pollution research international,2022,29(44):65993-66009. DOI: 10.1007/ S1135 6-022-21847-X.

[3] Linlin Z ,Jasper M ,Bill W , et al.Installation Quality Inspection for High Formwork Using Terrestrial Laser Scanning Technology [J].Symmetry,2022,14(2):377-377.DOI:10.3390/ SYM14020377.

[4] Jian L ,Yusen W ,Hongzheng L , et al.Pavement surface defect recognition method based on vehicle system vibration data and feedforward neural network[J].International Journal of Pavement Engineering, 2023,24(1): DOI:10.1080/ 10298436. 2023. 2188594.

[5] Wenzhi Y ,Qun Y .A Novel Method for Pavement Transverse Crack Detection Based on 2D Reconstruction of Vehicle Vibration Signal[J].KSCE Journal of Civil Engineering, 2023, 27 (7):2868-2881.DOI:10.1007/S12205-023-1972-1.

[6] Guido R ,Ferrisi S ,Lofaro D , et al.An Overview on the Advancements of Support Vector Machine Models in Healthcare Applications: A Review[J]. Information, 2024, 15 (4): 235-.DOI:10.3390/INFO15040235.

[7] Zhang W ,Yan S ,Tang H , et al.An improved shuffled frog leaping algorithm based on support vector machine for hybrid flow shop rescheduling with disturbance prediction[J].Swarm and Evolutionary Computation,2025,93101843-101843. DOI: 10.1016/J.SWEVO.2025.101843.

[8] Hui W, Guizhi W .Improving random forest algorithm by Lasso method[J].Journal of Statistical Computation and Simulation, 2021,91(2):353-367.DOI:10. 1080/ 00949655. 2020. 1814776.

[9] Wang Q ,Chen H .Optimization of parallel random forest algorithm based on distance weight[J].Journal of Intelligent & Fuzzy Systems,2020,39(2):1-13.DOI:10.3233/jifs-179965.

[10] Xingyu F ,Yingyue C ,Jingru Y , et al.BGRF: A broad granular random forest algorithm[J].Journal of Intelligent & Fuzzy Systems,2023,44(5):8103-8117.DOI:10.3233/JIFS-223960.

[11] Zhao L ,Zhang Z .A improved pooling method for convolutional neural networks.[J].Scientific reports,2024, 14 (1): 1589-1589.DOI:10.1038/S41598-024-51258-6.

[12] Chaoyang H ,Fengjiao W ,Chi A , et al.An Improved New Convolutional Neural Network Method for Inverting the Pore Pressure in Oil Reservoir by Surface Vertical Deformation [J]. LITHOSPHERE,2021,2021.DOI:10.2113/2021/5597238.

[13] Legnaro E ,Guastavino S ,Piana M , et al.Deep Learning for Active Region Classification: A Systematic Study from Convolutional Neural Networks to Vision Transformers[J].The Astrophysical Journal,2025,981(2):157-157.DOI:10. 3847/ 1538-4357/ADB41A.

[14] Abolfazl A ,Biswajeet P ,M. A A .An ensemble architecture of deep convolutional Segnet and Unet networks for building semantic segmentation from high-resolution aerial images[J]. Geocarto International,2022,37(12):3355-3370.DOI: 10.1080/ 10106049.2020.1856199.