A Novel Hybrid Feature Optimization Framework for Road Surface Identification with Vision Transformer and EGWO-SVM

Ramya Krishna Rajavolu
Department of EECE, GITAM School of Core Engineering, GITAM (Deemed to be University), Visakhapatnam, India
ramyakrishna.71@gmail.com (corresponding author)

Lakshmi Rajeswara Rao Langoju
Department of EECE, GITAM School of Core Engineering, GITAM (Deemed to be University), Visakhapatnam, India
rlangoju@gitam.edu

Received: 25 November 2025 | Revised: 16 December 2025 and 4 January 2026 | Accepted: 6 January 2026

 DOI: https://doi.org/10.48084/etasr.16508

ABSTRACT

Accurate identification of road surface conditions is essential for autonomous driving, Intelligent Transportation Systems (ITS), and driver-assist technologies. This paper proposes an integrated hybrid framework that combines multi-resolution texture decomposition, attention-based deep learning, handcrafted statistical descriptors, and evolutionary feature optimization. The approach begins by applying Stationary Wavelet Transform (SWT) to extract stable texture characteristics from road images. A Vision Transformer enhanced with Squeeze-and-Excitation (ViT-SE) blocks captures global spatial relationships, whereas Gray-Level Co-Occurrence Matrix (GLCM) descriptors contribute fine-grained statistical features. These complementary features are filtered using Mutual Information (MI) and optimized using Enhanced Grey Wolf Optimization (EGWO), enabling compact and discriminative feature selection. The optimized feature subset is classified using a Support Vector Machine (SVM). Experiments conducted on a multi-class road surface dataset demonstrate that the proposed SWT–ViT-SE–GLCM–MI–EGWO–SVM pipeline achieves 99.52% accuracy, outperforming state-of-the-art Convolutional Neural Network (CNN), Transformer, and hybrid models. The results confirm the synergy of multi-domain features and metaheuristic optimization in real-time road condition assessment applications.

Keywords-road surface classification; Vision Transformer (ViT); Stationary Wavelet Transform (SWT); Gray-Level Co-Occurrence Matrix (GLCM); Mutual Information (MI); Enhanced Grey Wolf Optimization (EGWO); Support Vector Machine (SVM)

I. INTRODUCTION

Reliable perception of road surface conditions is fundamental for Advanced Driver-Assistance Systems (ADAS), Intelligent Transportation Systems (ITS), and fully autonomous vehicles because variations in surface state—dry asphalt, wet pavement, compacted snow, gravel, and standing water—directly influence tire–road friction, braking distance, and vehicle stability [1, 2]. Misclassification of these conditions can lead to unsafe control actions, particularly at high speeds or under adverse weather. Figure 1 illustrates representative road surface images considered in this study, including snowy footprints, cracked asphalt, dry concrete, debris-covered surfaces, potholes filled with water, and highly reflective wet asphalt; these examples highlight the large intra-class variability and inter-class similarity that any robust classifier must handle. Recent research has increasingly adopted deep learning and Transformer-based architectures to address such challenges. Vision Transformers (ViTs) have been used for crack detection on asphalt and concrete, achieving superior robustness to noise and complex backgrounds compared with conventional Convolutional Neural Networks (CNNs) [1], whereas fast inference ViT variants such as LeViT have demonstrated efficient and accurate automatic pavement distress classification [2]. Hybrid and multimodal approaches further extend these ideas: multimodal Transformer models fuse road images with auxiliary time-series data to classify winter road surface conditions [3], and deep neural network–Support Vector Machine (SVM) hybrids integrating Gray-Level Co-Occurrence Matrix (GLCM), Histogram of Oriented Gradients (HOG), and ResNet-50 features have been shown to improve road surface condition identification across dry mud, fresh snow, and water-asphalt categories [4]. At the same time,

CNN–Transformer fusion networks for visual–tactile sensing exploit both camera images and intelligent tire measurements to enhance terrain recognition under varying illumination [5]. In the broader context of road and remote-sensing scene understanding, Swin-Transformer encoders and hybrid attention networks have proven highly effective for high-resolution road extraction and semantic parsing [6, 7]. More recently, contrastive-learning-based methods that first extract road areas and then classify their surface conditions have reported additional gains in classification accuracy and generalization [8]. Building on these advances, the present work focuses on a hybrid multi-domain framework that leverages global attention, multi-resolution texture information, and optimized feature selection to accurately distinguish the diverse road surface types exemplified in Figure 1.

Reliable road surface identification directly affects braking distance, tire–road friction, and vehicle stability, making it a crucial component of ADAS and autonomous vehicles. Although deep learning and Transformer-based models achieve promising accuracy, they often ignore multi-resolution texture cues and suffer from feature redundancy. This motivates a hybrid framework that integrates texture, global attention, and optimized feature selection.

Fig. 1. Road surface images – examples of snowy, cracked, debris-covered, pothole, and wet reflective surfaces.

II. RELATED WORK

Road surface analysis research falls into three categories: image-based categorization of surface types, deep and Transformer-based pavement fault and condition models, and hybrid feature engineering and optimization for robust classification. Classical texture descriptors and basic classifiers were used in early road surface studies. Authors in [9] evaluated road marking visibility for camera and LiDAR systems under varying weather conditions, highlighting the influence of surface material and reflectivity on machine vision performance. Authors in [10] used GLCM descriptors to characterize surface texture and traditional machine-vision techniques to distinguish road types, whereas authors in [11] used GLCM features with an LGBM classifier to automatically categorize damaged road surfaces from camera images. These studies demonstrate the discriminative capability of second-order texture statistics, but their use of hand-designed features hinders scalability for complicated, large-scale datasets.

Multimodal and sensor-fusion techniques have been studied more recently due to deep learning. Authors in [12] developed an artificial neural network that classifies pavement types using audio and picture data, boosting robustness when visual information is compromised by shadows or low illumination. Authors in [13] suggested a Transformer-based model that predicts paved and unpaved road conditions from onboard accelerometers and gyroscopes, enabling surface classification even when cameras are blocked. Authors in [3] developed a multimodal Transformer that blends road images with weather and traffic data to achieve state-of-the-art performance on dry, slushy, slippery, and snow-packed winter roads.

Pavement fracture and distress detection, although not focused on surface-type classification, provides useful insights into feature learning and model construction. Authors in [14] presented a comprehensive evaluation of machine-vision-based crack detection algorithms and highlighted the increasing relevance of CNNs and Transformers for managing noisy and heterogeneous pavement data. Authors in [15] suggested a Transformer-based crack segmentation network for embedded hardware that balances accuracy and computational cost using ViT architectures. Authors in [16] used CNNs trained on the CRACK500 dataset to automatically classify cracks, showing great performance but generalization issues. Crack-focused experiments show that attention-based models are better at capturing long-range contextual information on road surfaces, but they miss multi-resolution texture cues and suffer from feature redundancy.

Recent attention has therefore focused on hybrid deep–handcrafted systems that incorporate the strengths of both paradigms. Authors in [4] proposed a GHR50 model that fuses ResNet-50 features with GLCM and HOG descriptors, followed by an SVM classifier, achieving high accuracy across several road surface categories but at the cost of a high-dimensional feature space. Additional studies employed GLCM- and Local Binary Patterns (LBP)-based descriptors with k-Nearest Neighbors (kNN) or SVM classifiers to discriminate asphalt, gravel, and other surface types, supporting the importance of texture modeling. In parallel, wavelet-domain approaches have been employed to characterize pavement texture. Wavelet- and stationary-wavelet-based descriptors correlate with skid resistance and provide robust multi-scale texture representations in related surface inspection tasks.

Many hybrid techniques lack global attention modeling, which is typical of Transformer designs, or systematic feature selection mechanisms to reduce redundancy and computational cost. Thus, high-dimensional road surface analysis pipelines require effective feature selection and dimensionality reduction algorithms. Authors in [17] proposed a normalized Mutual Information (MI)-driven band grouping and minimum-redundancy–maximum-relevance scheme for hyperspectral feature extraction, which performed competitively with kernel SVM classifiers. MI has been widely used to measure feature relevance and redundancy. The Grey Wolf Optimizer (GWO) and its variations are powerful feature selection metaheuristics

for high-dimensional classification problems. In security and biomedical datasets, adaptive and multi-strategy GWO frameworks remove extraneous features and improve classifier accuracy [18-20].

These findings indicate that MI-based relevance filtering combined with advanced GWO-based feature selection can produce compact and highly discriminative feature subsets. To the best of our knowledge, no prior study has integrated Stationary Wavelet Transform (SWT), Vision Transformer enhanced with Squeeze-and-Excitation (ViT-SE) blocks, GLCM texture descriptors, MI-based ranking, and an Enhanced Grey Wolf Optimization (EGWO)–driven feature selection strategy within a unified road surface classification framework. To address this gap, the SWT–ViT-SE–GLCM–MI–EGWO–SVM pipeline uses multi-domain features and metaheuristic optimization to achieve high classification accuracy and feature efficiency.

III. METHODOLOGY

The proposed hybrid road surface classification framework integrates multi-resolution texture decomposition, attention-based deep feature extraction, handcrafted statistical descriptors, feature relevance evaluation, metaheuristic feature selection, and SVM-based classification. The complete design pipeline is illustrated in Figure 2, where the input road surface images undergo sequential processing beginning with preprocessing, followed by SWT, ViT-SE, GLCM feature extraction, MI scoring, EGWO, and final decision-making via SVM.

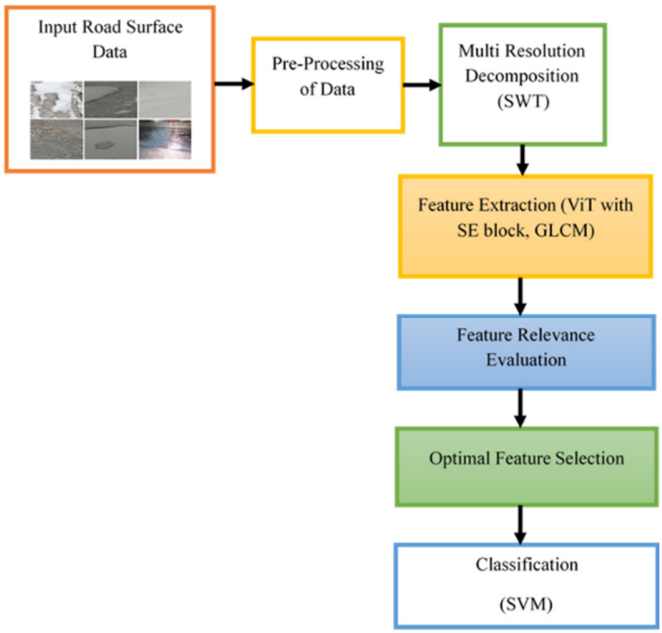

Fig. 2. Design flow of the proposed methodology.

Preprocessing of road surface images is performed to minimize variations in illumination, noise, and camera artifacts. Each input image is resized to a fixed resolution to maintain consistency and processed using a Gaussian filter, which reduces high-frequency noise and stabilizes the texture patterns essential for SWT and GLCM. The Gaussian transformation applied to the dataset is shown in Figure 3. Mathematically, resizing is performed according to:

$$I_{\text{resized}} = T(I) \tag{1}$$

and Gaussian smoothing follows:

$$G(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \tag{2}$$

whereas normalized pixel intensities are obtained by:

$$I_{\text{norm}} = \frac{I - \mu}{\sigma} \tag{3}$$

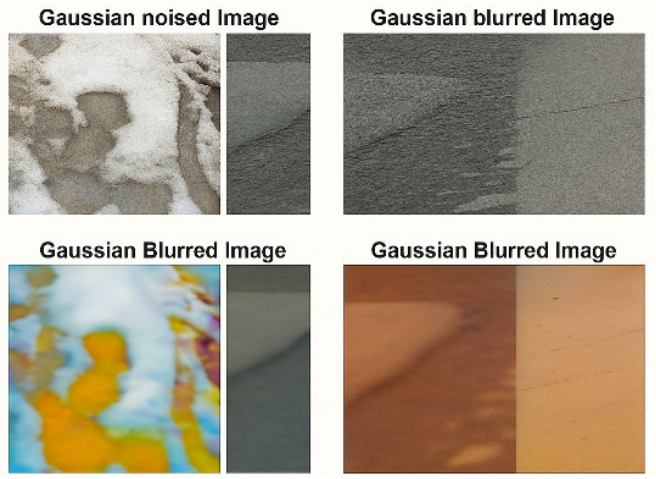

Fig. 3. Gaussian processed images.

After preprocessing, SWT decomposition is applied to obtain multi-resolution texture representations that are shift-invariant and robust to illumination changes. SWT decomposes the input image into approximation and detail coefficient sub-bands at different scales. The outputs of Level 1 and Level 2 SWT, comprising horizontal, vertical, diagonal, and approximation components, are depicted in Figure 4. These multi-scale coefficients preserve both global structure and fine-grained texture cues necessary for accurate road surface discrimination. SWT's ability to capture subtle irregularities has been emphasized in recent studies on multi-resolution image analysis and pavement texture modeling [21, 22].

Next, deep feature extraction is performed using a ViT-SE module. The ViT processes images by dividing them into sequences of fixed-size patches, enabling the model to learn long-range contextual relationships through multi-head self-attention. The complete processing flow—including input tokenization, positional encoding, and Transformer encoding—is illustrated in Figure 5. SE blocks further refine the extracted features by recalibrating channel-wise responses, using global average pooling followed by two fully connected layers. This enhances the discriminative power of the learned features, especially for visually similar surface types. Recent

advancements demonstrate that ViT-based models significantly outperform conventional CNNs for road and material surface classification due to their superior global attention mechanism [23, 24].

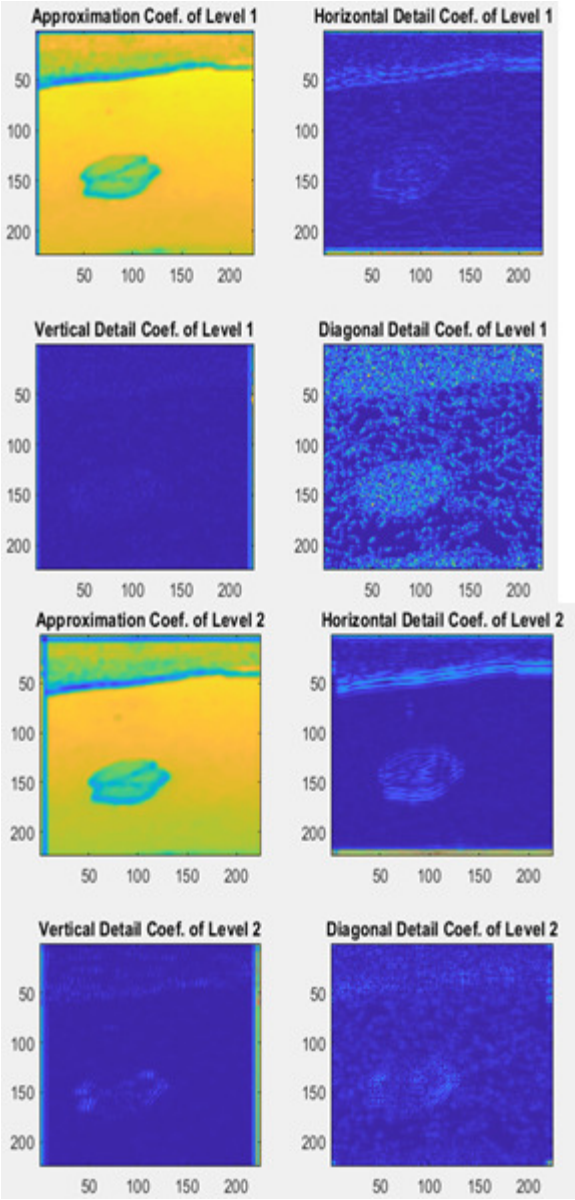

Fig. 4. Detailed H, V, D coefficient outputs of Level 1 and Level 2.

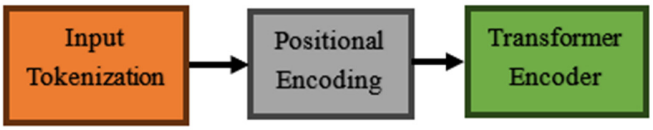

Fig. 5. Flow of the ViT.

To complement the Transformer-based global features, handcrafted texture descriptors are extracted using the GLCM, which captures second-order statistics such as contrast, correlation, energy, homogeneity, and entropy. The normalized co-occurrence matrix is defined as:

$$P(i,j) = \frac{C(i,j)}{\sum_{i,j} C(i,j)} \tag{4}$$

and texture measures are computed using:

$$\text{Contrast} = \sum_{i,j} (i-j)^2 P(i,j) \tag{5}$$

$$\text{Correlation} = \frac{\sum_{i,j} (i-\mu_i)(j-\mu_j) P(i,j)}{\sigma_i \sigma_j} \tag{6}$$

$$\text{Energy} = \sum_{i,j} P(i,j)^2 \tag{7}$$

$$\text{Homogeneity} = \sum_{i,j} \frac{P(i,j)}{1+(i-j)^2} \tag{8}$$

$$\text{Entropy} = -\sum_{i,j} P(i,j) \log(P(i,j)+\epsilon) \tag{9}$$

GLCM descriptors have been widely used for modeling subtle pavement textures such as cracks, roughness, or wetness [25]. Once both deep and handcrafted features are extracted, MI is used for relevance filtering, retaining only the most discriminative attributes. MI between a feature X and surface-class label Y is expressed as:

$$MI(X;Y) = \sum_x \sum_y p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \tag{10}$$

which eliminates redundant and non-informative features. This improves the efficiency of the subsequent optimization stage. To further refine the feature set, EGWO is employed. EGWO simulates the leadership hierarchy and hunting strategies of grey wolves—alpha, beta, and delta—to perform optimal feature subset selection. The wolf position updates follow:

$$D_\alpha = | C_1 X_\alpha - X |, \quad X_1 = X_\alpha - A_1 D_\alpha \tag{11}$$

with the final update rule:

$$X(t+1) = \frac{X_1 + X_2 + X_3}{3} \tag{12}$$

and feature-selection fitness evaluated as:

$$F = \alpha \cdot Acc + \beta \left(1 - \frac{|S|}{N}\right) \tag{13}$$

where $|S|$ is the selected feature count, ensuring both accuracy and compactness. Recent research confirms EGWO's superior performance in high-dimensional optimization tasks and its effectiveness in classification problems requiring minimal feature redundancy [26]. Finally, the optimized features are classified using an SVM equipped with a Radial Basis Function (RBF) kernel. SVMs are chosen for their excellent generalization ability in high-dimensional feature spaces and

their proven reliability in surface condition classification tasks. The combination of SWT, ViT-SE, GLCM, MI, EGWO, and SVM enables a powerful multi-domain learning framework capable of accurately identifying various road surface conditions under diverse environmental scenarios. An SVM is employed instead of the ViT Multi-Layer Perceptron (MLP) head due to its superior generalization performance on optimized hybrid feature sets and limited training samples.

IV. EXPERIMENTAL RESULTS

A. Experimental Setup

The performance of the proposed hybrid SWT–ViT-SE–GLCM–MI–EGWO–SVM framework was rigorously evaluated through a series of controlled experiments designed to assess both classification accuracy and computational efficiency. All experiments were conducted using a high-performance computing environment equipped with an Intel® Core™ i9-12900K processor, 32 GB of DDR5 RAM, an NVIDIA® RTX 3080 GPU, and running Windows 11. The complete system configuration is summarized in Table I, whereas the hyperparameters and model settings—including a batch size of 32, an Adam optimizer with a 1e-4 learning rate, a feature vector length of 1024 prior to optimization, an RBF kernel SVM classifier, and GWO settings of 20 wolves over 50 iterations—are listed in Table II. These settings were determined empirically to ensure optimal convergence and stability during training.

TABLE I. SPECIFICATIONS FOR IMPLEMENTATION

Device	Specification
Processor	Intel® Core™ i9-12900K CPU @ 3.20 GHz
RAM	32 GB DDR5
GPU	NVIDIA® GeForce RTX 3080 (10 GB VRAM)
Storage	1 TB NVMe SSD
Operating system	Windows 11 Pro

TABLE II. EXPERIMENTAL PARAMETERS AND SETTINGS

Parameter	Value
Training batch size	32
Learning rate (ViT fine-tuning)	1e-4
Optimizer	Adam
Feature vector length (pre-optimization)	1024
SVM kernel	RBF
GWO population size	20 wolves
Max iterations	50

A balanced multi-class road surface dataset with 200–500 images per class was used. Nevertheless, the model's dependence on static images may limit generalization under extreme real-world conditions.

Although the proposed framework achieves high classification accuracy, several measures were incorporated to prevent overfitting and ensure reliable generalization. First, a stratified data partitioning strategy was adopted, dividing the dataset into 70% training, 15% validation, and 15% testing subsets to avoid bias across surface classes. Second, MI-based relevance filtering and EGWO were applied to eliminate redundant and non-informative features, thereby reducing model complexity and variance. Additionally, an ablation study was conducted to verify that performance gains were incremental and consistent across pipeline stages. The use of an RBF kernel SVM further enhances generalization by controlling decision boundaries in high-dimensional feature spaces. The consistent results observed in confusion matrices and Receiver Operating Characteristic (ROC) curves across all classes confirm that the reported accuracy is not a result of overfitting.

B. Dataset Description

The experimental evaluation was conducted using a balanced real-world road surface image dataset comprising 27 surface categories. Each class contains approximately 200–500 labeled images captured under varying illumination, weather, and surface conditions, ensuring substantial intra-class variability. The dataset was curated from real traffic environments and manually verified for label consistency. A stratified sampling strategy was employed to preserve class balance during training, validation, and testing. While the dataset is not publicly hosted due to collection constraints, it can be made available to researchers upon reasonable request to support reproducibility and further comparative studies [27].

A stratified split of 70% training, 15% validation, and 15% testing was employed to guarantee unbiased evaluation across all surface types. Representative prediction outputs from the trained model are displayed in Figure 6, illustrating accurate visual classification across all surface categories.

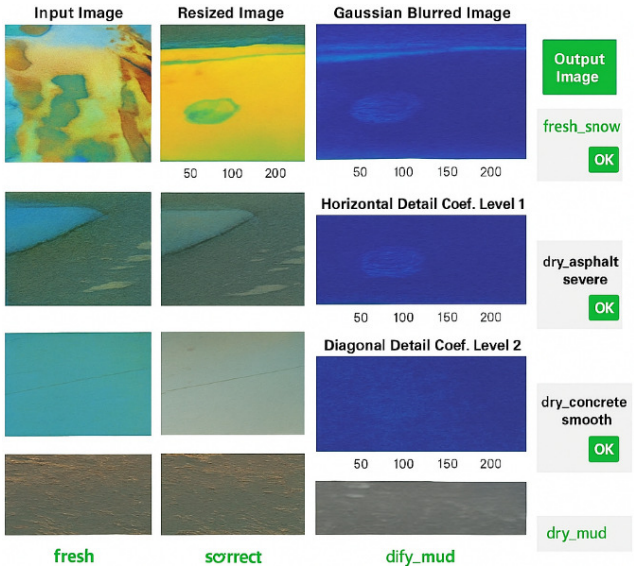

Fig. 6. Outputs achieved for the prediction of road surface conditions.

C. Performance Metrics

Performance analysis was conducted using key evaluation metrics: accuracy, precision, sensitivity, specificity, F1-score and the number of selected features. These metrics were computed at every stage of the feature extraction and optimization pipeline to illustrate the incremental benefits contributed by each component. The results are provided

comprehensively in Table III, which compares baseline configurations (GLCM-only, ViT-only, ViT+SE, and ViT+GLCM) with the final optimized model. The proposed method achieved exceptional performance, reaching 99.52% accuracy, 99.50% sensitivity, 99.49% specificity, 99.38% precision, and an F1-score of 99.44%, while reducing the feature dimensionality from 956 to only 512. This significant reduction in feature size highlights the impact of MI-based filtering and EGWO-driven optimization in improving both accuracy and computational efficiency.

TABLE III. PERFORMANCE METRICS OF BASELINE AND PROPOSED MODELS

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)
GLCM+SVM	87.52	87.75	86.78	87.25
ViT+SVM	91.78	91.76	91.72	91.50
ViT+SE+SVM	93.93	93.91	93.89	93.60
ViT+GLCM+SVM	97.80	97.73	97.70	97.60
Proposed (SWT–ViT-SE–GLCM–MI–EGWO–SVM)	99.52	99.50	99.49	99.38

A comparison with state-of-the-art road surface classification methods reveals that the proposed approach outperforms all competing techniques, including traditional SVM models, PSO-SVM, deep learning architectures, and recent Transformer-based models such as ViT-MLP and GHR50-SVM. The proposed SWT–ViT-SE–GLCM–MI–EGWO–SVM framework achieved the highest reported accuracy of 99.52% and an F1-score of 99.44%, significantly surpassing earlier methods with reported accuracies ranging from 88% to 97.80%.

D. Model Evaluation

To further examine the classification robustness, a confusion matrix was generated for the final model and is illustrated in Figure 7. The confusion matrix confirms consistent classification across all 27 classes with minimal misclassification, demonstrating high discriminative capability even for visually similar surface types such as wet versus dry asphalt. Additionally, the model's diagnostic performance was assessed using ROC analysis. As depicted in Figure 8, the ROC curves exhibit a steep rise toward the upper-left corner, confirming excellent separation between classes and an almost perfect true-positive versus false-positive trade-off. The similarity of ROC curves across classes indicates balanced data distribution and consistent discriminative performance of the optimized feature set.

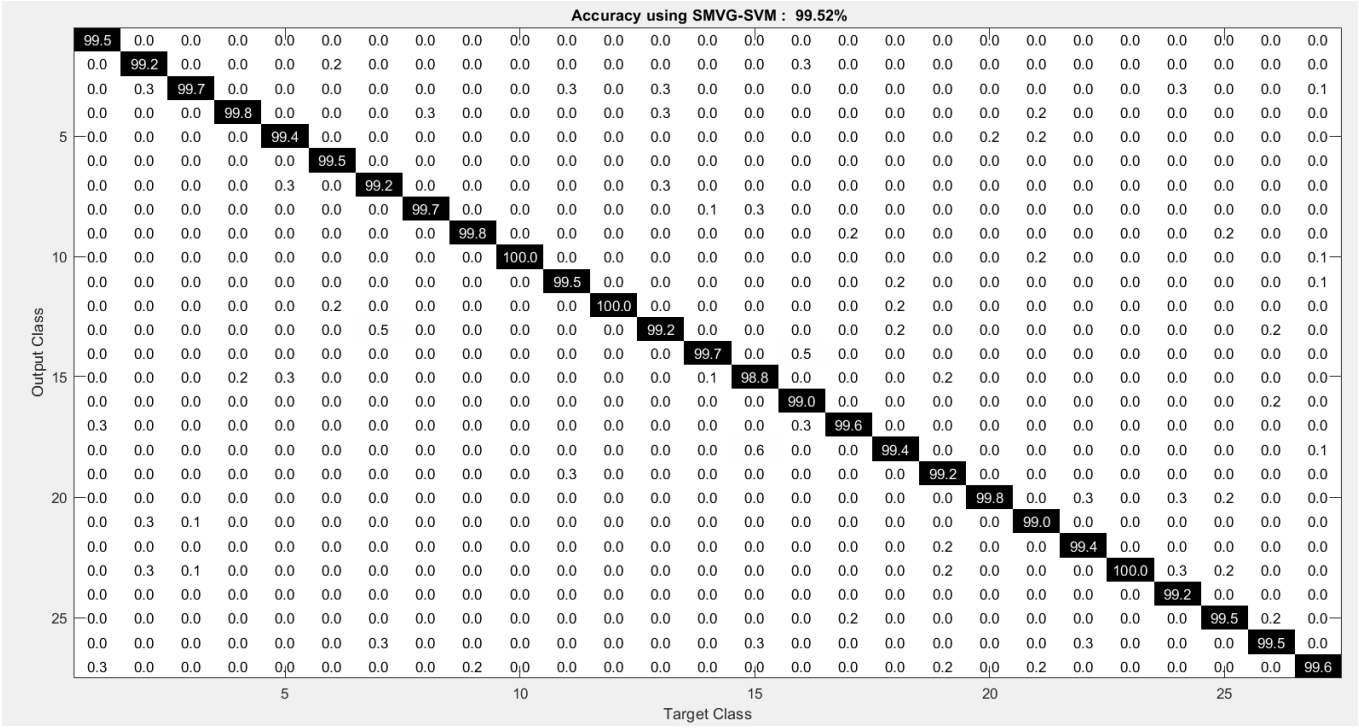

Fig. 7. Confusion Matrix of the proposed SWT–ViT-SE–GLCM–MI–EGWO–SVM framework.

E. Ablation Study

An ablation study was conducted to examine the individual effects of SE blocks, GLCM features, MI filtering, and GWO optimization. The findings, presented in Table IV, show clear performance improvements at each stage: adding SE blocks improved accuracy by approximately 1.95%; combining ViT and GLCM increased classification ability by capturing complementary global and texture features; MI filtering enhanced generalization by removing redundant descriptors; and GWO further reduced the feature set by nearly 45% without sacrificing accuracy. After EGWO, the reduced feature set of 512 dimensions achieved the peak accuracy of 99.52%, confirming the efficiency of the optimization process. Table IV

demonstrates that each component contributes incrementally to performance improvement, with EGWO yielding the most compact and accurate feature set.

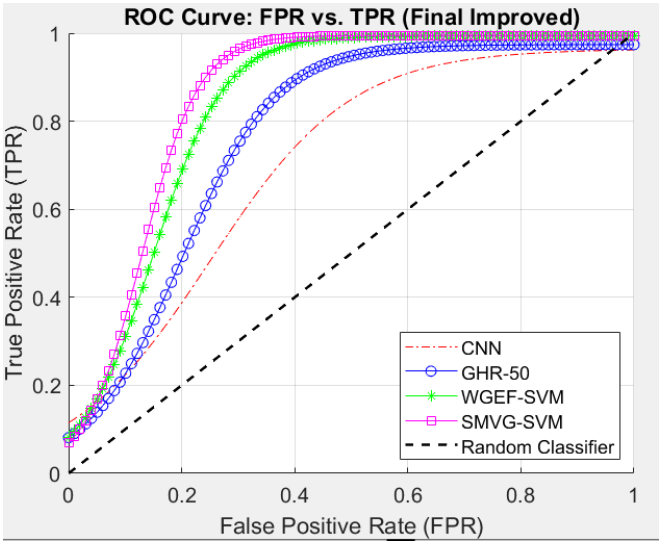

Fig. 8. ROC curves for the proposed and baseline models.

TABLE IV. FEATURE COUNT AND ACCURACY IN THE ABLATION STUDY

Feature selection stage	Number of features	Accuracy (%)
Before MI-GWO	956	94.65
After MI	726	97.48
After GWO	512	99.52

F. Limitations and Discussion

Although the results demonstrate outstanding performance, the study also acknowledges limitations. The reliance on 2D still images may reduce temporal continuity in rapidly changing surface conditions. The ViT component, while highly accurate, is computationally expensive during training, suggesting the need for lightweight transformer alternatives. Future enhancements could incorporate additional modalities such as thermal images or LiDAR, as well as Explainable Artificial Intelligence (XAI)-driven interpretability for safety-critical deployments. The experimental findings strongly validate that the proposed hybrid multi-domain learning framework achieves state-of-the-art performance in road surface classification, combining high accuracy, low feature dimensionality, and superior robustness across multiple real-world conditions. The integration of wavelet decomposition, Transformer-based attention, texture statistics, MI filtering, and EGWO optimization yields a highly reliable classification pipeline suited for ITS and autonomous vehicle platforms.

Although the proposed framework integrates multiple processing stages, its design carefully balances classification performance and computational efficiency. Computationally intensive components such as ViT feature extraction are applied only during offline training and feature generation. Feature dimensionality is significantly reduced from 956 to 512 using MI filtering and EGWO optimization, resulting in faster inference and lower memory requirements. The final SVM classifier operates on a compact optimized feature set, enabling real-time decision-making. Hardware specifications and execution parameters are reported to demonstrate feasibility on standard GPU-enabled systems. Thus, the framework maintains high accuracy while remaining practical for deployment in ITS and autonomous driving applications.

V. CONCLUSION

The proposed hybrid Stationary Wavelet Transform (SWT)–Vision Transformer enhanced with Squeeze-and-Excitation (ViT-SE)–Gray-Level Co-Occurrence Matrix (GLCM)–Mutual Information (MI)–Enhanced Grey Wolf Optimization (EGWO)–Support Vector Machine (SVM) framework demonstrates a powerful and efficient solution for automatic road surface classification. By combining multi-resolution wavelet decomposition, attention-enhanced Transformer features, and handcrafted texture descriptors, the model effectively captures both global and fine-grained surface characteristics. The integration of MI and EGWO significantly reduces feature redundancy while improving discriminative quality.

Experimental results confirm outstanding performance, achieving 99.52% accuracy, strong class separability, and excellent robustness across all road surface categories. In addition, the framework achieves high sensitivity, specificity, precision, and F1-score, highlighting its reliability for real-world applications. Overall, the system provides a reliable and computationally efficient approach suitable for Intelligent Transportation Systems (ITS), Advanced Driver-Assistance Systems (ADAS), and autonomous vehicle applications. Future work may explore lightweight Transformer variants, multimodal sensor fusion, and real-time deployment strategies to further enhance practical applicability.

While the proposed framework demonstrates strong performance on a balanced multi-class road surface dataset, evaluation on additional datasets would further strengthen generalization claims. Due to limited availability of publicly annotated road surface datasets with consistent labeling, cross-dataset testing was not conducted in the current study. However, robustness was examined through extensive ablation studies, class-wise confusion analysis, and Receiver Operating Characteristic (ROC) evaluation, all of which indicate stable performance across visually similar surface types. Future research will focus on validating the framework on larger and heterogeneous datasets, including cross-domain and multi-modal inputs, to further assess real-world deployment capability.

REFERENCES

[1] E. Asadi Shamsabadi, C. Xu, A. S. Rao, T. Nguyen, T. Ngo, and D. Dias-da-Costa, "Vision transformer-based autonomous crack detection on asphalt and concrete surfaces," *Automation in Construction*, vol. 140, Aug. 2022, Art. no. 104316, https://doi.org/10.1016/j.autcon.2022.104316.

[2] Y. Chen, X. Gu, Z. Liu, and J. Liang, "A Fast Inference Vision Transformer for Automatic Pavement Image Classification and Its Visual Interpretation Method," *Remote Sensing*, vol. 14, no. 8, Apr. 2022, Art. no. 1877, https://doi.org/10.3390/rs14081877.

[3] Y. Moroto, K. Maeda, R. Togo, T. Ogawa, and M. Haseyama, "Multimodal Transformer Model Using Time-Series Data to Classify Winter Road Surface Conditions," *Sensors*, vol. 24, no. 11, May 2024, Art. no. 3440, https://doi.org/10.3390/s24113440.

[4] R. R. Krishna and N. Jyothi, "Road Surface Condition Identification with Deep Neural Networks and SVM Classifier," *Engineering, Technology & Applied Science Research*, vol. 15, no. 2, pp. 21998–22003, Apr. 2025, https://doi.org/10.48084/etasr.10166.

[5] R. Shi *et al.*, "CNN-Transformer for visual-tactile fusion applied in road recognition of autonomous vehicles," *Pattern Recognition Letters*, vol. 166, pp. 200–208, Feb. 2023, https://doi.org/10.1016/j.patrec.2022.11.023.

[6] T. Chen, D. Jiang, and R. Li, "Swin Transformers Make Strong Contextual Encoders for VHR Image Road Extraction," in *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, Kuala Lumpur, Malaysia, 2022, pp. 3019–3022, https://doi.org/10.1109/IGARSS46834.2022.9883628.

[7] J. Li, Y. Zhang, P. Yun, G. Zhou, Q. Chen, and R. Fan, "RoadFormer: Duplex Transformer for RGB-Normal Semantic Road Scene Parsing," *IEEE Transactions on Intelligent Vehicles*, vol. 9, no. 7, pp. 5163–5172, July 2024, https://doi.org/10.1109/TIV.2024.3388726.

[8] W. H. Abdulsalam, R. H. Ali, S. H. Jadooa, and S. S. Hussein, "Automated Glaucoma Detection Techniques: A Literature Review," *Engineering, Technology & Applied Science Research*, vol. 15, no. 1, pp. 19891–19897, Feb. 2025, https://doi.org/10.48084/etasr.9316.

[9] T. E. Burghardt *et al.*, "Visibility of various road markings for machine vision," *Case Studies in Construction Materials*, vol. 15, Dec. 2021, Art. no. e00579, https://doi.org/10.1016/j.cscm.2021.e00579.

[10] S. Marianingsih, W. Widodo, M. S. S. Pieter, E. V. Manullang, and H. Y. Nanlohy, "Machine Vision for the Various Road Surface Type Classification Based on Texture Feature," *Journal of Mechanical Engineering Science and Technology*, vol. 6, no. 1, pp. 40–47, July 2022, https://doi.org/10.17977/um016v6i12022p040.

[11] Robet, C. Juliandy, Andi, Hendri, J. Hendrik, and F. A. Tarigan, "Image Road Surface Classification Based on GLCM Feature Using LGBM Classifier," *IOP Conference Series: Earth and Environmental Science*, vol. 1083, no. 1, Sept. 2022, Art. no. 012006, https://doi.org/10.1088/1755-1315/1083/1/012006.

[12] J. K. Lee, B. K. Kim, H. Choi, and S. I. Chang, "Road-pavement classification by artificial neural network model based on tire-pavement noise and road-surface image," *Applied Acoustics*, vol. 225, Nov. 2024, Art. no. 110194, https://doi.org/10.1016/j.apacoust.2024.110194.

[13] I. Aslam and S. Mahfuz, "Transformer-Based Classification of Road Conditions Using Vehicular Sensor Data," *Procedia Computer Science*, vol. 257, pp. 444–451, Jan. 2025, https://doi.org/10.1016/j.procs.2025.03.058.

[14] S. Huang, H. Chen, L. Yan, X. Zou, B. Li, and Y. Bi, "A review of the progress in machine vision-based crack detection and identification technology for asphalt pavements," *Digital Transportation and Safety*, vol. 4, no. 1, pp. 65–79, Mar. 2025, https://doi.org/10.48130/dts-0025-0006.

[15] Y. Zhu, T. Cao, and Y. Yang, "A Transformer-Based Pavement Crack Segmentation Model with Local Perception and Auxiliary Convolution Layers," *Electronics*, vol. 14, no. 14, July 2025, Art. no. 2834, https://doi.org/10.3390/electronics14142834.

[16] S. Matarneh, F. Elghaish, D. J. Edwards, F. P. Rahimian, E. Abdellatef, and O. Ejohwomu, "Automatic crack classification on asphalt pavement surfaces using convolutional neural networks and transfer learning," *Journal of Information Technology in Construction*, vol. 29, no. 55, pp. 1239–1256, Dec. 2024, https://doi.org/10.36680/j.itcon.2024.055.

[17] M. R. Islam, B. Ahmed, M. A. Hossain, and M. P. Uddin, "Mutual Information-Driven Feature Reduction for Hyperspectral Image Classification," *Sensors*, vol. 23, no. 2, Jan. 2023, Art. no. 657, https://doi.org/10.3390/s23020657.

[18] G. Li, Y. Cui, and J. Su, "Adaptive mechanism-based grey wolf optimizer for feature selection in high-dimensional classification," *Plos One*, vol. 20, no. 5, May 2025, Art. no. e0318903, https://doi.org/10.1371/journal.pone.0318903.

[19] Y. Wang, Y. Yin, H. Zhao, J. Liu, C. Xu, and W. Dong, "Grey wolf optimizer with self-repulsion strategy for feature selection," *Scientific Reports*, vol. 15, no. 1, Apr. 2025, Art. no. 12807, https://doi.org/10.1038/s41598-025-97224-8.

[20] A. Dede *et al.*, "Wavelet-Based Feature Extraction for Efficient High-Resolution Image Classification," *Engineering Reports*, vol. 7, no. 2, Feb. 2025, Art. no. e70027, https://doi.org/10.1002/eng2.70027.

[21] F. Li *et al.*, "2D-wavelet based micro and macro texture analysis for asphalt pavement under snow or ice condition," *Journal of Infrastructure Preservation and Resilience*, vol. 2, no. 1, May 2021, Art. no. 14, https://doi.org/10.1186/s43065-021-00029-y.

[22] Q. Cui, Y. Li, H. Bian, J. Kong, and Y. Dong, "Visual defect recognition with stationary wavelet transform based neural networks," *Digital Signal Processing*, vol. 158, Mar. 2025, Art. no. 104947, https://doi.org/10.1016/j.dsp.2024.104947.

[23] Y. Xu, C. Zhang, and H. Li, "Transformer-based large vision model for universal structural damage segmentation," *Automation in Construction*, vol. 176, Aug. 2025, Art. no. 106256, https://doi.org/10.1016/j.autcon.2025.106256.

[24] T. Ahmed, N. Ejaz, and S. Choudhury, "Redefining Real-Time Road Quality Analysis With Vision Transformers on Edge Devices," *IEEE Transactions on Artificial Intelligence*, vol. 5, no. 10, pp. 4972–4983, Oct. 2024, https://doi.org/10.1109/TAI.2024.3394797.

[25] M. H. Daneshvari, E. Nourmohammadi, M. Ameri, and B. Mojaradi, "Efficient LBP-GLCM texture analysis for asphalt pavement raveling detection using eXtreme Gradient Boost," *Construction and Building Materials*, vol. 401, Oct. 2023, Art. no. 132731, https://doi.org/10.1016/j.conbuildmat.2023.132731.

[26] M. Yu, J. Xu, W. Liang, Y. Qiu, S. Bao, and L. Tang, "Improved multi-strategy adaptive Grey Wolf Optimization for practical engineering applications and high-dimensional problem solving," *Artificial Intelligence Review*, vol. 57, no. 10, Sept. 2024, Art. no. 277, https://doi.org/10.1007/s10462-024-10821-3.

[27] T. Zhao and Y. Wei, "A road surface image dataset with detailed annotations for driving assistance applications," *Data in Brief*, vol. 43, Aug. 2022, Art. no. 108483, https://doi.org/10.1016/j.dib.2022.108483.