International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Transfer Learning-based Ensemble Classifier for Multi-Class Texture Classification in Remote Sensing

G. Sujatha[1*], Syeda Sanobar Ali[2], Patimeeni Shiva Mukesh[2], Alval Abhilash Gou[2]
[1]Assistant Professor, [2]UG Student, [1,2]Department of Computer Science and Engineering (AI & ML),
[1,2]Kommuri Pratap Reddy Institute of Technology, Ghanpur, Ghatkesar, 501301, Telangana, India.
*Correspondence: G. Sujatha (gsujatha39@gmail.com)

Abstract
Automated texture classification has become an important area of research in computer vision and industrial inspection systems, where accurate surface analysis is required to maintain product quality and detect defects during manufacturing. In earlier industrial practices, surface inspection was performed manually by human experts who visually examined materials to identify defects such as holes, cracks, and foreign objects. Although manual inspection was widely used, it was often slow, inconsistent, and prone to human error. Later, traditional automated systems based on classical image processing techniques were introduced, but these methods relied heavily on handcrafted features and simple classification approaches, which limited their ability to capture complex texture patterns. As industrial datasets became larger and more diverse, these limitations created a need for intelligent systems capable of extracting meaningful visual features and performing reliable multi-class classification. To address these challenges, this study presents a transfer learning–based texture classification framework that combines deep feature extraction with machine learning techniques. In the proposed approach, a pretrained Visual Geometry Group 19 (VGG19) is used to extract high-level visual features from industrial texture images. The extracted feature vectors are then used to train multiple machine learning classifiers including Linear Discriminant Analysis Classifier (LDAC), Quadratic Discriminant Analysis Classifier (QDAC), Support Vector Classifier (SVC), and Extra Trees Classifier (ETC). Among these models, the proposed VGG19 with ETC classifier achieved the highest classification accuracy of 97.31%, demonstrating its strong capability in distinguishing between different texture categories such as good, hole, and objects.

Keywords: Texture Classification, Industrial Inspection, Computer Vision, Transfer Learning, Deep Feature Extraction, Visual Geometry Group 19 (VGG19).

1. Introduction

Remote sensing (RS) is the science of collecting information about objects without any direct physical contact, typically through a satellite, aircraft or unmanned aerial vehicle (UAV). Examples of applications of remote sensing include geological survey, environment testing, oil exploration, traffic management, earthquake prediction, and water conservancy construction. Remote-sensing images have improved in both spatial and temporal resolutions with the evolution of satellite sensors, which provides opportunities in resolving fine details on the earth's surface as shown in Figure 1. Satellites such as MODIS (1 km × 1 km) offering thermal data with high temporal resolution suffer from low spatial resolution. Landsat, on the other hand, offers small-scale variations of 100–200 m but with very low temporal resolution. The new generation of satellites can deliver very high spectral and spatial images; for example, IKONOS-2 generates images with 4-band multispectral resolution and spatial resolution from 2.5 to 4 m. Unmanned aerial vehicles (UAVs) present an improved solution of remote-sensing acquisition platforms, which witnessed a high level of growth in past years and are used widely for fire detection, surveillance mapping, and landslide monitoring, among other uses [2]. UAVs have several advantages over satellite and aerial images. First, they are easier to deploy to satisfy the requirements of rapid monitoring, assessment, and mapping. They can work at lower altitudes compared to the piloted aircraft, which provides spatial resolution at the centimeters level.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

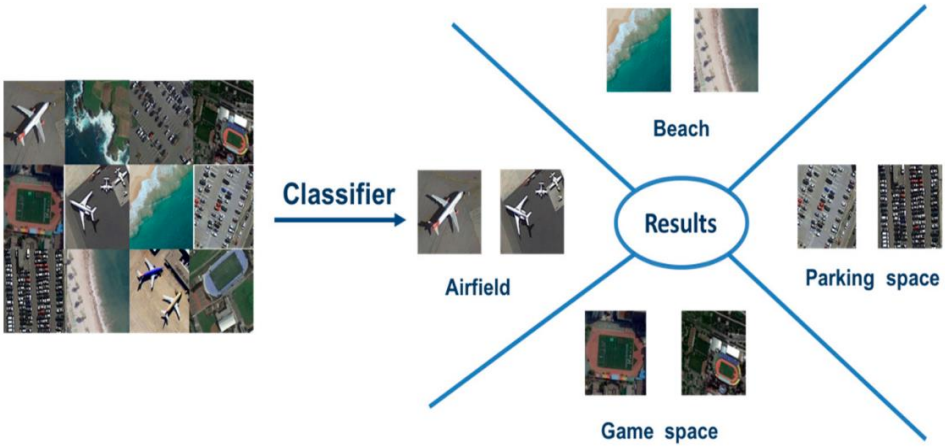

Figure. 1: Remote sensing scene classification.

They can fly any time the weather permits, leading to improvements in temporal resolution. As the spatial resolution increases, images are likely to contain noisy and outlying descriptors. A recent study [3] proposed a convolutional neural network (CNN) to classify images captured by a camera mounted on a UAV. CNN model to classify a digital surface model beside UAV images [4]. combined CNNs with object-based image analysis (OBIA) for land cover classification using Multiview data. The two-branch neural network to assign multiple class labels to UAV imagery [5].

2. Literature Survey

Bazi, et al. [6] proposed a remote-sensing scene-classification method based on vision transformers. These types of networks, which are now recognized as state-of-the-art models in natural language processing, do not rely on convolution layers as in standard convolutional neural networks (CNNs). Instead, they use multihead attention mechanisms as the main building block to derive long-range contextual relation between pixels in images. In a first step, the images under analysis are divided into patches, then converted to sequence by flattening and embedding. Carrilho, et al. [7] reviewed, they summarized the most important advancements in the last 5 years and focus mostly on machine learning-based approaches. They also outline the most promising avenues of research in the future.

Wang, et al. [8] developed and found that transformers need more parameters than CNNs. Additionally, further research is also needed regarding inference speed to improve transformers' performance. It was determined that the most common application scenes for transformers in our database are urban, farmland, and water bodies. They also found that transformers are employed in the natural sciences such as agriculture and environmental protection rather than the humanities or economics. Finally, this work summarizes the analysis results of transformers in remote sensing obtained during the research process and provides a perspective on future directions of development. Wang, et al. [9] introduced the fundamental concepts of transformers and highlight the first successful Vision Transformer (ViT). Building on the ViT, they reviewed subsequent improvements and optimizations introduced for image classification tasks. They then compare the strengths and limitations of these transformer-based models against classic CNNs through experiments. Finally, they explored key challenges and potential future directions for image classification transformers.

Aleissaee, et al. [10] reviewed the remote sensing community and has also witnessed an increased exploration of vision transformers for a diverse set of tasks. Although several surveys have focused on transformers in computer vision in general, to the best of their knowledge they are the first to present a systematic review of recent advances based on transformers in remote sensing. Their survey covers more than 60 recent transformer-based methods for different remote sensing problems in sub-areas of

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

remote sensing: very high-resolution (VHR), hyperspectral (HSI) and synthetic aperture radar (SAR) imagery. They concluded the survey by discussing different challenges and open issues of transformers in remote sensing. Tombe, et al. [11] presented a comprehensive review of the developments of various computer vision methods in remote sensing. There is currently an increase of remote sensing datasets with diverse scene semantics; this renders computer vision methods challenging to characterize the scene images for accurate scene classification effectively. This paper presented technology breakthroughs in deep learning and discusses their artificial intelligence open-source software implementation framework capabilities. Further, this paper discusses the open gaps/opportunities that need to be addressed by remote sensing communities.

Mao, et al. [12] presented this work comprehensively comparing the entire YOLO family, highlighting key innovations and their practical implications. They also discuss the challenges, including dataset limitations, domain generalization, and computational constraints, proposing future solutions such as synthetic data generation, federated learning, and edge AI deployment. By bridging the gap between academic advancements and industrial applications, this review is a practical guide for selecting and optimizing YOLO models for fabric inspection, paving the way for intelligent quality control systems. Zhang, et al. [13] proposed a new remote sensing scene classification method, Remote Sensing Transformer (TRS), a powerful "pure CNNs → Convolution + Transformer → pure Transformers" structure. First, they integrate self-attention into ResNet in a novel way, using our proposed Multi-Head Self-Attention layer instead of 3 × 3 spatial revolutions in the bottleneck. Then they connect multiple pure Transformer encoders to further improve the representation learning performance completely depending on attention. Finally, use a linear classifier for classification. They train our model on four public remote sensing scene datasets: UC-Merced, AID, NWPU-RESISC45, and OPTIMAL-31. The experimental results show that TRS exceeds the state-of-the-art methods and achieves higher accuracy.

Li, et al. [14] investigated unexplored ideas for remote sensing image captioning task, using a novel patch-level region-aware module with a multi-label framework. Due to an overhead perspective and a significantly larger scale in RSIs, a patch-level region-aware module is designed to filter the redundant information in the RSI scene, which benefits the Transformer-based decoder by attaining improved image perception. Technically, the trainable multi-label classifier capitalizes on semantic features as supplementary to the region-aware features. Moreover, modeling the inner relations of inputs is essential for understanding the RSI. Zhang, et al. [15] proposed a multiple hierarchical cross-scale Transformer model that efficiently combines the Transformer model with CNNs and is specifically designed for complex remote sensing scene classification. Firstly, a feature pyramid network with attention aggregation extracts the multi-scale base features. Then, these base features are fed into the proposed multi-scale channel Transformer (MSCT) module to derive the global features with channel-wise attention. Additionally, the base features are also fed into the proposed hierarchical cross-scale Transformer (HCST) module, which can obtain multi-level cross-scale representations.

Lu, et al. [16] proposed a novel aerial object detection framework called DFCformer. DFCformer is mainly composed of three parts: the backbone network DMViT, which introduces deformation patch embedding and multi-scale adaptive self-attention to capture sufficient features of the objects; FRGC guides feature interaction layer by layer to break the barriers between feature layers and improve the information discrimination and processing ability of multi-scale critical features; CAIM adopts an attention mechanism to fuse multi-scale features to perform hierarchical reasoning on the relationship

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

between different levels and fully utilize the complementary information in multi-scale features. Extensive experiments have been conducted on the FAIR1M dataset, and DFC former shows its advantages by achieving the highest scores with stronger scene adaptability.

3. Proposed Methodology

The proposed methodology follows a structured framework for developing an intelligent image analysis system capable of performing multi-class texture classification in industrial environments shown in figure 2. The methodology begins with dataset acquisition and image preprocessing, followed by deep feature extraction using transfer learning techniques. These extracted features are then utilized by multiple machine learning classifiers to analyses surface textures and identify different defect categories. The system integrates classical classification models with deep feature extraction to improve classification accuracy and reliability. A graphical interface allows users to interact with the system, perform dataset uploads, execute model training, and view classification results. This structured workflow ensures efficient data handling, model evaluation, and prediction generation for automated texture analysis.

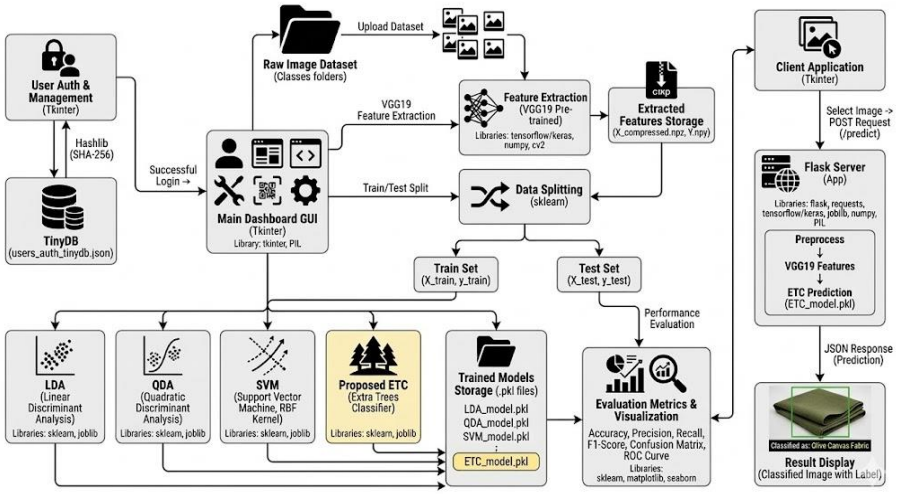

Figure. 2: System architecture of patch-level texture classification.

User Interface (Tkinter GUI)

- The user interacts with the system through a custom graphical interface developed using Tkinter.
- The interface facilitates critical workflow stages, including dataset uploading, feature extraction, model training, and performance evaluation.
- Users can visualize real-time classification results, confusion matrices, and detailed performance metrics.
- All GUI-triggered actions are communicated to backend modules that orchestrate the data analysis pipeline.

Dataset Input (Image Dataset)

- The system utilizes a structured dataset of industrial texture images categorized into specific class folders.
- These categories represent surface conditions such as normal surfaces, holes, or foreign objects.
- Images are loaded into the system and prepared for the preprocessing pipeline to ensure consistency across the dataset.

Image Preprocessing

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

- To ensure uniform dimensions for the neural network, all uploaded images are resized to a standard resolution.
- Images are converted into multidimensional arrays and normalized to optimize the efficiency of the feature extraction process.
- This standardized data format is then forwarded to the deep learning feature extractor.

Feature Extraction using VGG19

- A pretrained VGG19 convolutional neural network serves as the backbone for capturing high-level visual patterns.
- The final fully connected layers are truncated, allowing the convolutional base to act as a pure feature extractor.
- **Global Average Pooling** is applied to the output to generate compact numerical feature vectors that represent structural details and spatial patterns.

Train–Test Data Splitting

- The extracted feature vectors are partitioned into distinct training and testing subsets.
- The training set allows the classifiers to learn specific texture patterns, while the testing set provides an unbiased evaluation of the models' predictive power.
- This separation is crucial for validating the generalizability of the system on unseen industrial data.

Existing Baseline Models (LDAC, QDAC, SVC)

- The feature vectors are processed by several classical statistical and machine learning classifiers to establish a performance benchmark:
  - **LDAC:** Focuses on maximizing the linear separability between different surface classes.
  - **QDAC:** Models are more complex, quadratic decision boundaries for higher-dimensional data.
  - **SVC:** Utilizes kernel functions to identify the optimal hyperplane that separates texture categories.

Proposed Model (ETC)

- ETC is implemented as the core ensemble learning engine for the framework.
- It operates by constructing many randomized decision trees and aggregating their results to reach a final classification.
- By using randomized feature selection during tree construction, the model effectively reduces overfitting and improves the overall accuracy of texture classification.

Performance Evaluation and Result Analysis

- The outputs of both the baseline and proposed models are rigorously analyzed using a comprehensive suite of metrics.
- Evaluation includes Accuracy, Precision, Recall, and F1-score, alongside visual tools like Confusion Matrices and ROC Curves.
- This analysis identifies the most reliable classifier for specific industrial environments and surface types.

Prediction and Output Visualization

- In the final stage, the trained classifier is deployed to identify the class of new, unseen industrial images.
- New inputs pass through the same preprocessing and VGG19 feature extraction pipeline.
- The predicted class label is overlaid on the image and displayed within the Tkinter UI, allowing for rapid identification of surface conditions and defects.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

3.1 VGG19

The feature extraction stage plays a crucial role in transforming raw image data into meaningful representations that can be used for classification. In this study, a pretrained deep convolutional neural network is used to extract high-level visual features from industrial texture images. The network processes the input images through multiple convolutional and pooling layers to capture important spatial patterns such as edges, shapes, and texture structures. Instead of using the network for direct classification, the final classification layers are removed so that the deep network functions as a powerful feature extractor, as shown in fig 3. The resulting feature vectors represent the visual characteristics of each image and are later used by machine learning classifiers for texture classification.

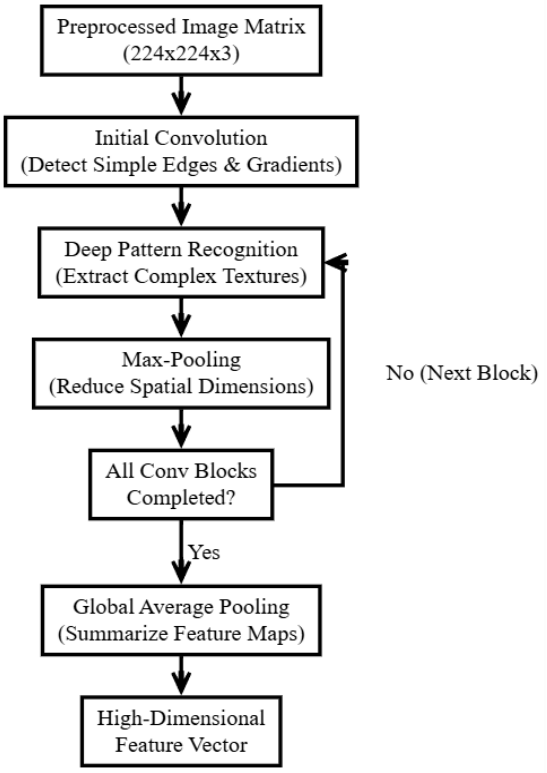

Figure. 3: Internal workflow of VGG19.

Input Image Preparation: The preprocessed images are provided as input to the deep convolutional network. Each image is resized to the required dimensions and formed into a numerical array structure. This standardized input ensures compatibility with the feature extraction model.

Convolutional Feature Learning: The input image passes through multiple convolutional layers within the network. These layers apply learnable filters that capture low-level visual patterns such as edges and color variations. As the layers progress, the network begins to identify more complex structures and texture patterns.

Activation and Non-Linear Transformation: After convolution operations, activation functions introduce non-linear transformations to the extracted features. This process enables the network to learn complex visual relationships within the image. The activation layers enhance the network's ability to represent intricate texture variations.

Spatial Down sampling through Pooling: Pooling layers are used to reduce the spatial dimensions of the feature maps generated by convolutional layers. This operation helps retain important visual information while reducing computational complexity. Pooling also improves the robustness of the feature representation against small image variations.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Feature Vector Generation: After the convolution and pooling operations are completed, global average pooling is applied to convert the feature maps into a compact numerical feature vector. This feature vector summarizes the most important visual information extracted from the image. The generated features are then forwarded to machine learning classifiers for the classification stage.

4. Result Description

The results of the study demonstrate the effectiveness of the proposed texture classification framework in analyzing industrial surface images. The system processes the input dataset through preprocessing, deep feature extraction, and machine learning classification stages to generate accurate predictions. Experimental results are obtained by training multiple classifiers and evaluating their performance using standard evaluation metrics. The results provide insights into how effectively the classifiers distinguish between different texture categories such as normal surfaces, holes, and foreign objects. Performance comparisons among the implemented models highlight the strengths and limitations of each classifier. These outcomes help determine the most reliable model for accurate industrial texture classification.

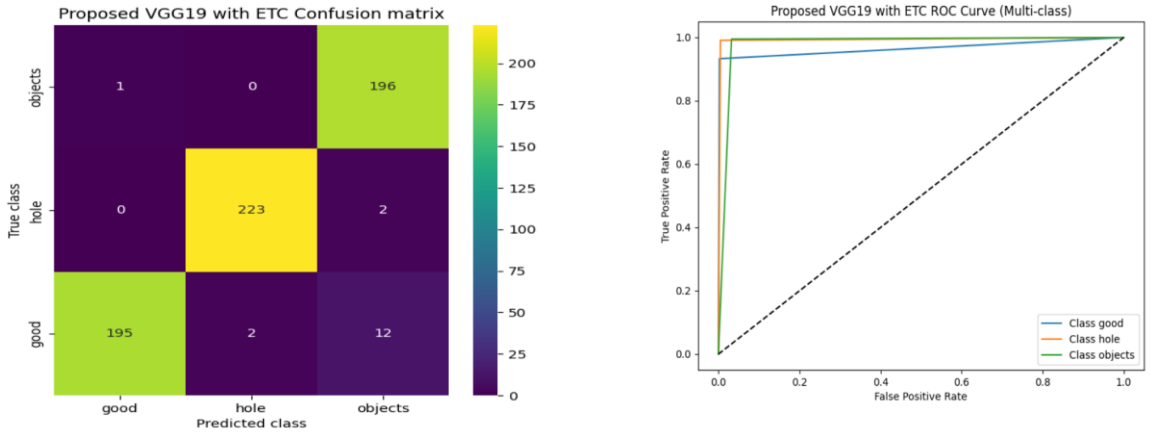

Figure 4: Performance Evaluation of Existing ETC Model (a) Confusion Matrix, (b) Multi-Class ROC Curve.

Figure 4(a) illustrates the confusion matrix obtained for the proposed classification approach that integrates VGG19 feature extraction with the ETC. The confusion matrix represents the relationship between the actual texture categories and the predicted categories generated by the classifier. It provides detailed information about the correct classifications and the misclassifications among the texture classes such as good, hole, and objects. This representation highlights the ability of the proposed model to correctly distinguish between different surface texture patterns present in the dataset. The distribution of values in the matrix indicates improved classification accuracy compared to the baseline models. This evaluation helps demonstrate the effectiveness of combining deep feature extraction with an ensemble learning classifier for multi-class texture classification.

Figure 4(b) depicts the multi-class ROC curve generated for the proposed VGG19 with ETC model. The ROC curve illustrates the relationship between the true positive rate and the false positive rate for each texture class across different classification thresholds. Separate curves are plotted for each category to evaluate how effectively the proposed model discriminates between the classes. The curves approaching the upper-left region of the graph indicate strong classification capability and improved predictive performance. This graphical representation highlights the robustness of the proposed approach in identifying texture categories with higher accuracy. The ROC analysis complements the confusion matrix by providing a visual assessment of the classifier's discriminative performance across multiple classes.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

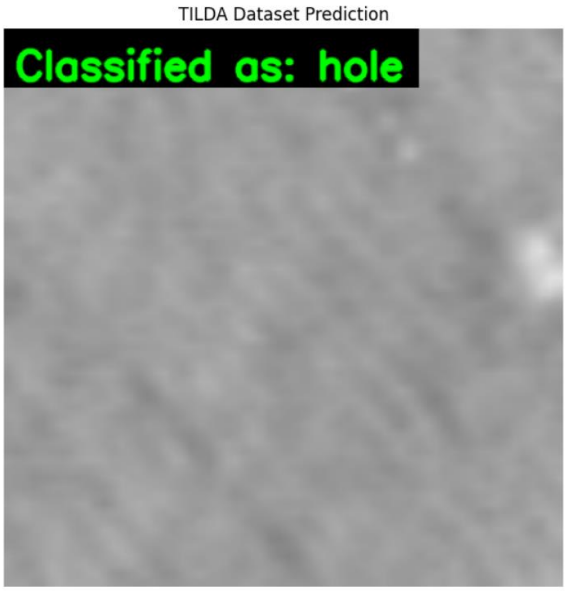

Figure. 5: Prediction result classified as Hole

Figure 5 illustrates the final prediction result obtained from the trained texture classification framework using a test image from the TILDA dataset. During this stage, the input image is processed through the same preprocessing and deep feature extraction pipeline using the VGG19 model. The extracted feature vector is then provided to the trained classifier, specifically the ETC, which determines the most appropriate texture category for the given image. The predicted output is displayed directly on the image to clearly indicate the classification result generated by the system. This visual representation confirms the ability of the trained model to analyze unseen images and correctly identify the texture class. The prediction stage demonstrates the practical applicability of the developed framework for automated texture recognition and surface defect identification.

Figure 6 illustrates another prediction result generated by the trained texture classification framework using a test image from the TILDA dataset. In this stage, the input image is processed through the preprocessing pipeline followed by deep feature extraction using the VGG19 convolutional neural network. The extracted feature vector is then provided to the trained ETC, which analyses the visual patterns present in the image to determine the appropriate texture category. The predicted class label is displayed directly on the image to clearly indicate the classification outcome. This result demonstrates the capability of the trained model to identify different texture conditions present in unseen images. The prediction output confirms the effectiveness of the integrated feature extraction and machine learning classification approach for automated texture recognition.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

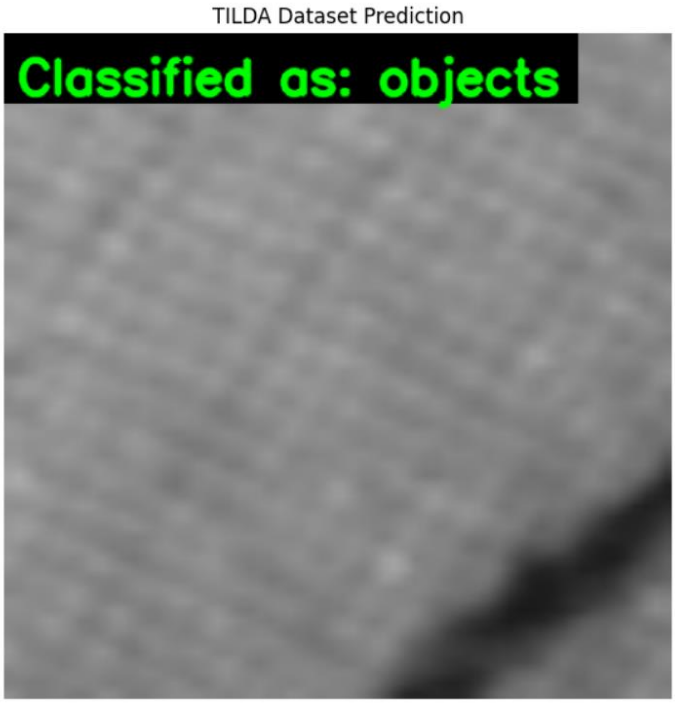

Figure. 6: Prediction result classified as Objects

The comparative analysis results presented in the table 1 highlight the performance differences among the implemented machine learning classifiers. The LDAC achieved an accuracy of 90.97%, demonstrating strong performance in classifying the texture categories with balanced precision, recall, and F-score values. The QDAC produced a comparatively lower accuracy of 48.34%, indicating that this model struggled to effectively distinguish between the texture classes in the dataset. The SVC showed moderate performance with an accuracy of 75.44%, providing better classification capability than QDAC but still lower than LDAC. In contrast, the proposed ETC achieved the highest accuracy of 97.31%, outperforming all other models in terms of precision, recall, and F-score. The results indicate that the ensemble learning capability of ETC enables more accurate identification of complex texture patterns.

Table 1**:** Comparative Performance Analysis of Machine Learning Classifiers for Texture Classification.

Model	Accuracy (%)	Precision (%)	Recall (%)	F-Score (%)
LDAC	90.97	91.10	91.19	90.87
QDAC	48.34	50.39	63.68	41.80
SVC	75.44	75.80	76.31	75.17
ETC	97.31	97.31	97.30	97.24

5. Conclusion

This research presents an intelligent texture classification framework for analyzing industrial surface images using transfer learning and machine learning techniques. The system integrates VGG19-based deep feature extraction with multiple classifiers including LDAC, QDAC, SVC, ETC. Experimental results demonstrate that the proposed approach significantly improves classification performance compared to traditional classifiers. Among the evaluated models, LDAC achieved an accuracy of 90.97%, SVC obtained 75.44%, and QDAC produced 48.34% accuracy. The proposed VGG19 with ETC model achieved the highest classification accuracy of 97.31%, along with improved precision,

recall, and F-score values. The use of deep feature extraction combined with an ensemble classifier enables more effective identification of texture categories such as good, hole, and objects. The experimental results confirm that the proposed framework provides reliable and accurate texture classification suitable for automated industrial inspection applications.

References

[1] Hu, Q.; Wu, W.; Xia, T.; Yu, Q.; Yang, P.; Li, Z.; Song, Q. Exploring the use of google earth imagery and object-based methods in land use/cover mapping. Remote Sens. 2013, 5, 6026–6042.

[2] Toth, C.; Jóźków, G. Remote sensing platforms and sensors: A survey. ISPRS J. Photogramm. Remote Sens. 2016, 115, 22–36.

[3] Hoogendoorn, S.P.; Van Zuylen, H.J.; Schreuder, M.; Gorte, B.; Vosselman, G. Microscopic traffic data collection by remote sensing. Transp. Res. Rec. 2003, 1855, 121–128.

[4] Valavanis, K.P. Advances in Unmanned Aerial Vehicles: State of the Art and the Road to Autonomy; Springer Science & Business Media: Berlin, Germany, 2008; ISBN 978-1-4020-6114-1.

[5] Sheppard, C.; Rahnemoonfar, M. Real-time scene understanding for UAV imagery based on deep convolutional neural networks. In Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, USA, 23–28 July 2017; pp. 2243–2246.

[6] Bazi, Y.; Bashmal, L.; Rahhal, M.M.A.; Dayil, R.A.; Ajlan, N.A. Vision Transformers for Remote Sensing Image Classification. Remote Sens. 2021, 13, 516. https://doi.org/10.3390/rs13030516.

[7] Carrilho, R.; Yaghoubi, E.; Lindo, J.; Hambarde, K.; Proença, H. Toward Automated Fabric Defect Detection: A Survey of Recent Computer Vision Approaches. Electronics 2024, 13, 3728. https://doi.org/10.3390/electronics13183728.

[8] Wang, R.; Ma, L.; He, G.; Johnson, B.A.; Yan, Z.; Chang, M.; Liang, Y. Transformers for Remote Sensing: A Systematic Review and Analysis. Sensors 2024, 24, 3495. https://doi.org/10.3390/s24113495.

[9] Wang, Y.; Deng, Y.; Zheng, Y.; Chattopadhyay, P.; Wang, L. Vision Transformers for Image Classification: A Comparative Survey. Technologies 2025, 13, 32. https://doi.org/10.3390/technologies13010032.

[10] Aleissaee, A.A.; Kumar, A.; Anwer, R.M.; Khan, S.; Cholakkal, H.; Xia, G.-S.; Khan, F.S. Transformers in Remote Sensing: A Survey. Remote Sens. 2023, 15, 1860. https://doi.org/10.3390/rs15071860.

[11] Tombe, R.; Viriri, S. Remote Sensing Image Scene Classification: Advances and Open Challenges. Geomatics 2023, 3, 137-155. https://doi.org/10.3390/geomatics3010007.

[12] Mao, M.; Hong, M. YOLO Object Detection for Real-Time Fabric Defect Inspection in the Textile Industry: A Review of YOLOv1 to YOLOv11. Sensors 2025, 25, 2270. https://doi.org/10.3390/s25072270.

[13] Zhang, J.; Zhao, H.; Li, J. TRS: Transformers for Remote Sensing Scene Classification. Remote Sens. 2021, 13, 4143. https://doi.org/10.3390/rs13204143.

[14] Li, Y.; Zhang, X.; Zhang, T.; Wang, G.; Wang, X.; Li, S. A Patch-Level Region-Aware Module with a Multi-Label Framework for Remote Sensing Image Captioning. Remote Sens. 2024, 16, 3987. https://doi.org/10.3390/rs16213987.

[15] Zhang, D.; Ma, W.; Jiao, L.; Liu, X.; Yang, Y.; Liu, F. Multiple Hierarchical Cross-Scale Transformer for Remote Sensing Scene Classification. Remote Sens. 2025, 17, 42. https://doi.org/10.3390/rs17010042.

[16] Lu, G.; He, X.; Wang, Q.; Shao, F.; Wang, H.; Wang, J. A Novel Multi-Scale Transformer for Object Detection in Aerial Scenes. Drones 2022, 6, 188. https://doi.org/10.3390/drones6080188.