IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 16 August 2024, accepted 17 September 2024, date of publication 23 September 2024, date of current version 9 October 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3466220

RESEARCH ARTICLE

Remote Sensing Image Interpretation: Deep Belief Networks for Multi-Object Analysis

MUHAMMAD WAQAS AHMED[1], ABDULLAH ALSHAHRANI[2], ABRAR ALMJALLY[3], NAIF AL MUDAWI[4], ASAAD ALGARNI[5], KHALED AL NOWAISER[6], AHMAD JALAL[1,7], AND JEONGMIN PARK[8]

[1]Department of Computer Science, Air University, Islamabad 44000, Pakistan
[2]Department of Computer Science and Artificial Intelligence, College of Computer Science and Engineering, University of Jeddah, Jeddah 21493, Saudi Arabia
[3]Department of Information Technology, College of Computer and Information Sciences, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh 13318, Saudi Arabia
[4]Department of Computer Science, College of Computer Science and Information System, Najran University, Najran 55461, Saudi Arabia
[5]Department of Computer Sciences, Faculty of Computing and Information Technology, Northern Border University, Rafha 91911, Saudi Arabia
[6]Department of Computer Engineering, College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia
[7]Department of Computer Science and Engineering, College of Informatics, Korea University, Seoul 02841, South Korea
[8]Department of Computer Engineering, Tech University of Korea, Siheung-si, Gyeonggi-do 15073, South Korea

Corresponding author: Abrar Almjally (aamjally@imamu.edu.sa)

ABSTRACT Object Classification in Remote Sensing Imagery holds paramount importance for extracting meaningful insights from complex aerial scenes. Conventional methods encounter challenges in achieving precision amidst diverse landscape features. This paper introduces an innovative hybrid model to enhance the accuracy of remote sensing multi-object classification. Incorporating a feature-level fusion approach inspired by successful methods, we leverage Adoptive Fuzzy C-means segmentation for precise object classification and Conditional Random Field labeling. Our model excels in capturing diverse features within remote sensing images using multiple feature extraction methods. The distinctive feature of our methodology lies in the thoughtful incorporation of a Deep Belief Network. Through rigorous experimental evaluations on two standard datasets, our proposed system demonstrates exceptional performance, emphasizing its significant potential for advancing methodologies in remote sensing multi-object classification. This tactful integration results in substantial improvements, yielding high accuracies of 97.24% (UCM) and 96.84% (RESISC45). The proposed model is methodologically novel and effective solution for advancing remote sensing image classification.

INDEX TERMS Multi object classification, remote sensing, feature fusion, object detection, deep belief network.

I. INTRODUCTION

Recent developments in imaging technology have elevated the resolution n of remote sensing (RS) imagery which makes it an essential tool for various research domains. This improved resolution enables enhanced capabilities, such as accurate object categorization, detailed change detection analysis, and comprehensive environmental monitoring. Aerial images, compared to their natural counterparts, vary significantly in size, orientation, and imaging environments [1], [2]. These variations pose challenges for accurate classification, impacting applications like urban construction, land-use classification, airport security and vegetation mapping. As imaging technology advances, the complexities of aerial images become more evident [3], [4]. Unlike terrestrial images, aerial images exhibit diverse interclass variability, where objects of the same type may appear in different sizes and orientations [5], [6]. Additionally, identical objects can change due to varying imaging conditions, such as equipment height during capture and solar altitude. These nuances make achieving precise classification a formidable task, crucial for societal development [7], [8], [9], [10].

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Wei.

The historical trajectory of remote-sensing image classification underscores the evolution from traditional low-level feature methods, such as Scale-Invariant Feature Transform, Color Histogram, Histogram of Oriented Gradients, and Local Binary Patterns, to more sophisticated approaches based on mid-level and high-level features [11], [12], [13], [14], [15]. The complexity inherent in modern remote-sensing images necessitated this shift, as the internal information within these images became more intricate. While mid-level features sought to capture global features by encoding extracted features, challenges persisted, particularly in terms of adaptability to different tasks and datasets [6], [16], [17], [18].

Aerial images, with their bird's eye view perspective, present difficulties in capturing spatial structure information and key features. The random distribution of key objects, coupled with the large shooting height and angle of aerial images, adds another layer of complexity to the understanding of these images. Due to the presence of useless background information, illumination changes and the angle of view makes accurate classification more difficult [19], [20], [21], [22].

Therefore, a new method for aerial image classification should be developed to address these obstacles. In this paper a model is proposed to tackle these obstacles in which a feature fusion method is presented to extract the informative features of the images prior to be fed into Deep-Belief Network for the classification and help overcoming the complexity of aerial images. This will be achieved by improving the local key feature representation, reducing computing complexity, and preserving features from different tiers. The main contribution of this papers are.

- We applied a combination of Gaussian Mixture Model and Adaptive Fuzzy C-Means for segmentation.
- Using Conditional Random Field as the post-processing and labeling algorithm, we handled the drawbacks related to segmentation that cannot addressed by other techniques and to classify the scene correctly.
- We fused wavelet packet features, Haar wavelet, Gabor and Correlogram, to improve classification accuracy.
- We modified Deep Belief Network for the final multi object classification.

The rest of the paper is organized as follows: Section II is about related works. In Section III, the methodology will be described that includes segmentation, labeling, feature extraction, and their fusion. Part 4 is about datasets, experimental design with results. Lastly, in Section V, the conclusions of the study are discussed.

II. RELATED WORKS

Image segmentation is one of the main task in analysis of aerial- derived remote sensing images (RS). The accuracy of segmentation method is crucial to feature extraction and classification in remote sensing images. Image segmentation has applications in many fields, each of which uses it in a different way. This diversity has contributed to the formulation of nuanced dynamics and metrics, shaping our strategic approach.

A. MULTI-OBJECT SEGMENTATION

In remote sensing the objective of the segmentation is to divide images into coherent regions. Recent advancements in Very High-Resolution remote sensing images include the introduction of a precise Mask Region Convolutional Neural Network as proposed by Wu et al [23]. It generates accurate segmentation masks along with bounding boxes for each instance. Unlike traditional Regions of Interest align methods, the proposed precise ROI pooling prevents accuracy loss. Semantic segmentation, which allocates homogeneous regions to distinct geographical object categories, has seen diverse approaches, including Markov random field models, level sets, clustering, and deep learning. Traditional techniques like clustering are less effective for high-resolution remote sensing images, because individual pixels within an object may possess distinct appearances. An optimized multi-kernel method designed for semantic segmentation utilizes an advanced Markov Random Field model for high spatial resolution remote sensing imagery [24]. Leveraging spatial and spectral data, the proposed Frequency Domain Feature-Guided Network improves semantic segmentation of Remote Sensing Images. FFGNet generates frequency domain characteristics by means of 2D discrete cosine transformation and patch partitioning. These properties are selectively amplified by the Frequency Enhancement Attention module, which then integrates with Spatial-Spectral Attention for enhanced spectral information. Up sampling and feature fusion highlight spectral subtleties in inference, a new loss function combines frequency and cross-entropy losses. FFGNet shows better performance on LoveDA and ISPRS Potsdam datasets [25].

N. Zhou et al. [26] suggests a novel MsASNet for the purpose of categorizing landslide data. This network, which is built upon the U-Net architecture has several distinctive characteristics. Here, one essential upgrade is the assimilation of a dynamic visual receptor that improves the network's ability to extract features from the given data. Further, a convolutional block attention module which will enhance information fusion and increase the accuracy of image segmentation is used. By collecting both local and global contexts, the proposed Geometric Prior-Guided Interactive Network solves the semantic segmentation in high-resolution remote sensing images. GPINet uses Local-Global Interaction Modules with dual-branch encoder to couple CNN and transformer outputs through cross-attention, so refining features. A new Geometric Prior Generation Module uses geometric priors iteratively to guide feature recovery. By weighted summing, the upsampled decoded features are finally merged with these geometric priors, hence improving pixel-level semantic correctness. On benchmark datasets, GPINet shows better performance than others, therefore confirming the efficiency of its geometric priors and model size control [27].

K. Chen et al. [28] put forward an improved method RSPrompter for remote sensing images, which improves SAM by introducing semantic category knowledge related to semantic category derived from prompt learning. This enhancement is conducive to SAM to output the semantic distinguishable segmentations for the RS images and therefore make a further expand its applicability and application performance in this line. Sulaiman and Isa [29]. introduce adoptive fuzzy K-means clustering for image segmentation, allowing multiple levels of membership. Fuzzy logic enables data members to be simultaneously allocated to multiple clusters, especially relevant in remote sensing images. Jia et al. [30]. propose the Hierarchical Heterogeneous Graph method, effective for change detection which is very difficult in radar images due to high Nosie which is overcome by using object based coarse to fine change segmentation. Ghadi et al. [31]. contribute methods for generating pre-source feature representations, enhancing precision in multi-source area networks. Zhang et al. [32]. propose an isotropic spatial energy function for class co-occurrence relationships in aerial-based remote sensing imagery, presenting an alternative to contemporary techniques. Li et al. [33] proposed transformer based model AAFormer for semantic segmentation for remote sensing images. An attention gate is utilized to refine the self-attention module to focus on the informative features. The model introduce the lightweight attention attend transformer block which helped to attain the contextual information. Rahaman et al. [34] research focuses on the impact of simulated labels and noise on urban water body image segmentation using deep learning algorithms with Sentinel-2 satellites. They utilized the U-Net architecture, mainly created for deep auto encoder-type networks with skip connections, as the main segmentation model. They create different training datasets in the simulated label noise that are used to evaluate how it affects the U-Net performance. The kind of synthetic noise considered includes salt and pepper errors with the help of Gaussian noise and also image translation for registration errors whereas less frequently observed noise types like mirroring are also considered. The state of the art fusion-based segmentation models has outperformed the traditional single-modal based methods in the case of semantic segmentation of remote sensing data. However, these models are usually based on CNNs or the ViT for fusion, which causes constraints in the local-global contextual modeling and representative aspects. To overcome these problems, certain solutions have been suggested, for example the use of both CNNs and transformers, to combine the advantages of both architectures. In particular, some multilevel fusion approaches have been investigated for shallow and deep features to improve feature extraction of semantic information and spatial context [35]. Li et al. [36] proposed a novel approach for pixel level segmentation for remote sensing images based on spectrum space Collaborative network. In order to enhance the feature representation, it employs a joint spectral–spatial attention mechanism that integrates spectral attention and spatial attention. To capture the spectral context, it uses Euclidean distance and for capturing the spatial context it uses position wise self-attention.

B. MULTI OBJECT CLASSIFICATION

In the domain of object categorization, researchers face challenges such as object localization, analyzing relationships, handling occluded components, and achieving effective class separation. Over the past decade, the bag-of-features model has been a dominant paradigm for imagery categorization [37]. Martin et al. [38]. introduced a Bayesian inference model for object tracking by enhancing object recognition precision and convergence rates. In a unique class-specific illustration technique utilized Gaussian mixture models and Euclidean distances for object categorization [39]. Another approach focused on multi-object categorization for indoor-outdoor scenes, employing segmentation and multiple kernel learning [40]. Wong et al. [41]. proposed a kernel learning approach for online object detection and classification, demonstrating effectiveness in rapidly tracking objects. Sumbul et al. [42]. developed methods incorporating a multi-source region network for object location representations, utilizing multispectral approaches for improved accuracy. To overcome the problem, such as low accuracy and difficulty of integrating multimodal features, of conventional methodologies, Zhang [43] puts forward a new way for the recognition and classification of multimodal remote sensing data. The proposed model uses a heat map and HGR correlation pooling fusion operation to integrate these two operations. It presents a new HGR correlation pooling fusion algorithm that successfully restores the initial signal while maintaining the transmitted information with the highest accuracy. The implemented algorithm improves feature learning for images by effectively employing multimodal information with different values of relevance. The integration of these components into a single approach gives a complete solution for raising the correctness and performance of identifying and categorizing multimodal remote sensing data.

Object classification challenges include localization, analyzing connections, recovering hidden features, and achieving desired outcomes. Ahmed et al. [2] developed a model that involves Fuzzy C-Means segmentation and saliency mapping as a means for identifying attention regions with greater focus in images. Thus, it generates a more detailed differentiation of the image areas that are important for feature extraction, which ultimately lead to higher levels of classify cation accuracy. Bo and Sminchisescu [44]defined characteristics using Gaussian mixture models, employing Euclidean distances for image comparisons. A. Ahmed et al [45]. utilized multiple kernel learning for multi-object categorization, enhancing classification with area-specific signatures. Ansith et al. [46]. presented a modified GAN architecture for land usage classification in high-resolution remote sensing images showcasing better results than most of the other deep learning methods. Huang et al. [47] recommended

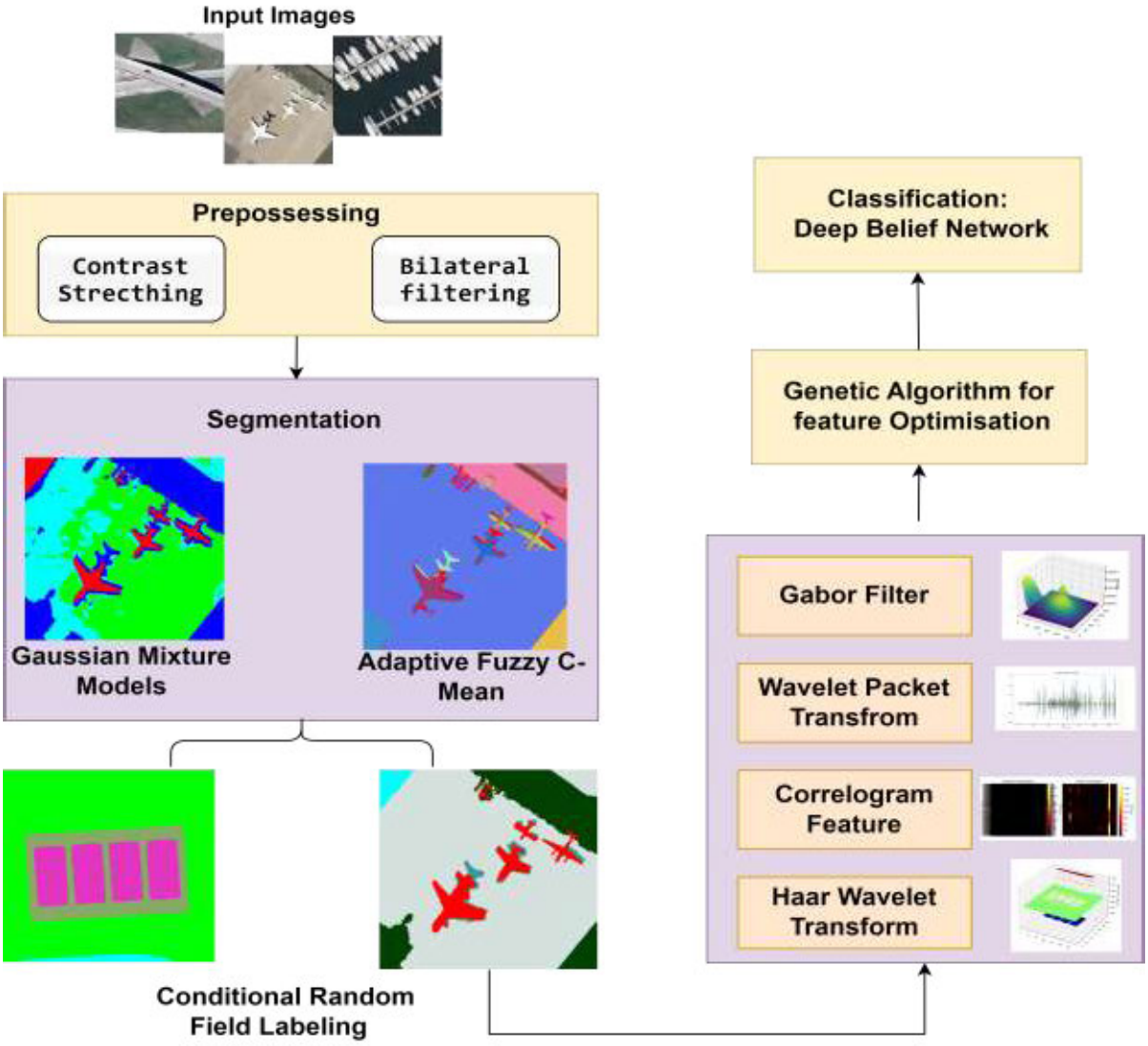


FIGURE 1. Proposed model architecture illustrating the key novel components and methodological flow.

SVM and PCA for high-resolution image classification, focusing on detecting object contours and motion. Object classification across different high-resolution images including Multispectral, Hyperspectral and Multi-temporal images a major challenge due to integration of both spectral and spatial domains. Several past approaches are weak in capturing multiple features necessary for classification of patients into different subgroups. To tackle this, Zheng et al. [48] introduced the Hybrid Fusion Net model that integrates 2D & 3D convolutional neural networks & a transformer encoder. This composed architecture enables HFN to compute multi-dimensional features in the spectral, spatial, and temporal domains while passing the learned global saliency and discriminative information to the transformer encoder.

III. PROPOSED MODEL

Our proposed model presents a holistic and intricate approach tailored for the precise classification of aerial images. During preprocessing stage, we used contrast stretching and bilateral filtering techniques to elevate image quality by mitigating noise and enhancing overall clarity. Our segmentation model integrates Gaussian Mixture Model and Adaptive Fuzzy C-Means however for further processing AFCM is considered due to the fact that it produced greater accuracy with better computational time. The subsequent stage involves feature extraction, where our model leverages a diverse set of techniques, including Gabor, Haar wavelet, wavelet packet transform, and Correlogram which helped in contributing to a robust feature set. The extracted features are then concatenated to facilitate fusion, paving the way for a more comprehensive representation of the image characteristics. Genetic algorithm is used for feature optimization as it fine-tunes the concatenated feature set which enhances the discriminative power of the model. To cap off this intricate framework, we adopted a Deep Belief Network for the final classification task. The integration of these preprocessing,

segmentation, feature extraction, optimization, and classification techniques underscores the model's efficacy and versatility, positioning it as a comprehensive and effective solution for aerial image classification in diverse scenarios. Fig. 1 display the structural diagram of proposed model.

A. DATA PROCESSING

In the initial stages of image classification for aerial imagery, our preprocessing pipeline contains three main step image resizing, contrast stretching and bilateral filtering. In order to make all the images consistent from both datasets we resized all the images to 213 X 213 pixels. Although the image size in both the datasets is 256 X 256 pixels but they exhibit significant variations in ground sampling distance and geographical coverage. We used bilinear interpolation for the resizing process as it provides good balance between the quality of image and the computational efficiency. The output pixel values are calculated using the weighted average of 2 × 2 neighborhood of the pixel surrounding the input coordinate which helps to maintain edge sharpness. The resizing of the images also helped in normalizing the scale variations in the datasets as same objects have different size in different classes and also in same classes. Contrast stretching is employed to expand the dynamic range of pixel values in the image, enhancing its overall clarity. This technique operates through a linear scaling transformation, effectively adjusting pixel intensities to span a wider range. The process is parameterized by minimum and maximum intensity values, as well as new minimum and maximum values, ensuring a controlled enhancement without introducing exaggeration [49], [50]. Mathematically, this can be expressed as:

$$O(x, y) = (I(x, y) - \min Val)/(maVal - minVal) \times (newMax - newMin) + newMin \quad (1)$$

where $O(x,y)$ is the output pixel intensity, $I(x,y)$ is the input pixel intensity, and $minVal$, $maxVal$, $newMin$, and $newMax$ are the parameters controlling the stretching. Additionally, bilateral filtering is implemented as a non-linear smoothing mechanism to mitigate noise while preserving essential image features, especially edges [51], [52]. The bilateral filter considers both spatial and intensity variations in the image, with a spatial Gaussian kernel addressing spatial differences and a range Gaussian kernel accounting for intensity disparities [53], [54]. Mathematically, the bilateral filter can be represented as:

$$\mathrm{O}\,(\mathrm{x},\mathrm{y}) = \frac{1}{\mathrm{w}} \sum_{i=-k}^{k} \sum_{j=-k}^{k} I\,(x+i, y+j)\,.\exp\left(\frac{i^2+j^2}{2\sigma_s^2}\right) .\exp\left(-\frac{(I\,(x+i, y+j) - I\,(x, y))^2}{2\sigma_\gamma^2}\right) \quad (2)$$

where $O(x, y)$ is the output pixel intensity, $I(x, y)$ is the input pixel intensity, σ_s is the spatial Gaussian kernel, σ_r is the range Gaussian kernel, and W is the normalization factor. This combined preprocessing approach serves to optimize the visual quality of the images by making them more conducive to subsequent image classification tasks by ensuring improved feature visibility and minimizing unwanted noise [55], [56].

FIGURE 2. Aerial images after application of bilateral filtering.

B. SEGMENTATION

With regard to aerial image segmentation, clustering algorithms are crucial in dividing an image into meaningful regions. So in this paper two different techniques Gaussian Mixture Models and Adaptive Fuzzy C-Means are used due to the fact that different character of aerial images makes certain approaches relevant in capturing individual patterns.

The computational time and segmentation results are evaluated after applying GMM and AFCM to the aerial images. The results showed that AFCM has faster computation times along with better segmentation accuracy. Due to the fuzzy membership values it is able to handle the inherent variability which is present in aerial images which is the base for accurate and smoother segmentations.

1) GMM SEGMENTATION

Gaussian Mixture Models segmentation is a probabilistic approach which is used in image processing for partitioning an image into distinct regions based on the assumption that the image is a mixture of several Gaussian distributions. The GMM models the pixel intensities of an image as a combination of these Gaussian components. The segmentation process assigs each pixel to the most likely Gaussian component by determining the region to which it belongs [57]. The GMM probability density function for a pixel intensity x is given by:

$$P\,(x) = \sum_{i=1}^{K} \pi_i . N(x|\mu_i, \sigma_i^2) \quad (3)$$

where k is the number of Gaussian components, π_i is the ith component, $N(x|\mu_i, \sigma_i^2)$ is the Gaussian distribution with mean μ_i and variance σ_i^2. It assigs each pixel to the Gaussian component with the highest probability. The parameters π_i, μ_i, σ_i^2 of the GMM are estimated using the expectation-Maximization algorithm. The EM algorithm iteratively refines the estimates to maximize the likelihood of the observed data [57]. Although GMM segmentation can capture complex data distributions, it is essential to consider

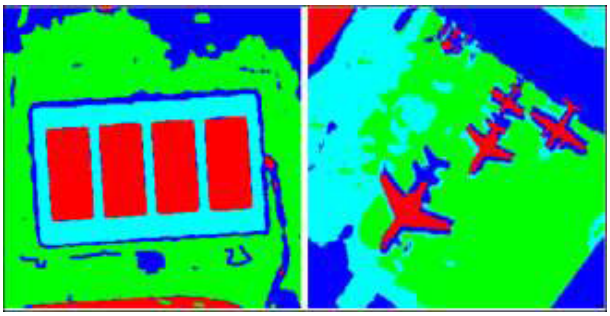


FIGURE 3. Segmented images using gaussian mixture model.

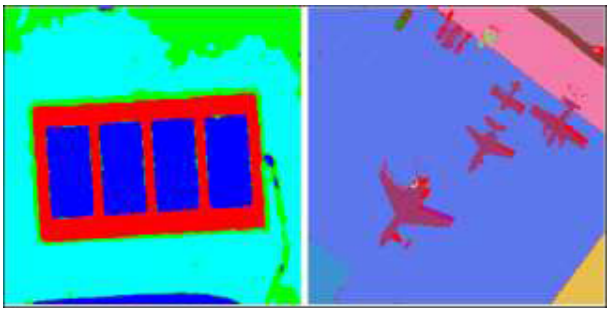


FIGURE 4. Segmented images using adoptive fuzzy C mean.

computational efficiency, especially for large-scale datasets. In scenarios where computational time is a critical factor, alternative methods like Adaptive Fuzzy C-Means may be considered due to their faster processing capabilities.

2) ADOPTIVE FUZZY C MEAN SEGMENTATION

Adaptive Fuzzy C-Means segmentation is suited for scenarios where the data has more complex and exhibit more changes. AFCM uses the Fuzzy C-Means algorithm by incorporating adaptive spatial information into the clustering process due to which it is able to handle non-uniform regions along with varying noise levels better than GMM [58], [59]. The AFCM objective function is formulated as follows:

$$J = \sum_{i=1}^{c} \sum_{j=1}^{n} \mu_{ij}^{m} . d_{ij}^{2} \tag{4}$$

where J represents the objective function to be minimized, c is the number of clusters, n is the total number of pixels, μ_{ij} is the membership value for pixel j belonging to cluster I and m is the fuzziness parameter controlling the degree of fuzziness. d_{ij} is the distance between the feature vector of pixel j and the centroid of cluster i [60]. The membership values are updated iteratively using the following formula:

$$\mu_{IJ} = \left[\sum_{k=1}^{c} \left[\left(\frac{d_{ij}}{d_{kj}} \pi \right) \right]^{\frac{2}{m-1}} \right]^{-1} \tag{5}$$

The spatial information is carried out with the help of adaptive spatial constraint by modifying the distance term d_{ij} based on the local spatial structure of the image. This adaptive term helps AFCM to be more robust in the presence of varying textures and spatial non-uniformities. The AFCM algorithm iteratively refines the membership values and cluster centroids until it converges as a result it provides a segmented image where each pixel is assigned membership values for multiple clusters. The final segmentation is achieved by assigning each pixel to the cluster with the highest membership value. At time AFCM can show unsatisfactory results if noise is too much in the images so it is important that we use some filter during preprocessing stage in order to eliminate the noise so that AFCM can perform better as in this model we used adoptive mean filter in preprocessing stage. The result of few of the segmented images after applying Adoptive Fuzzy C Mean which in display also depicts better results than GMM which is displayed in Fig. 4.

Upon comparing the time estimates and segmentation accuracy of two methods, Adaptive Fuzzy C-Mean segmentation was chosen for labeling due to its faster processing time compared to GMM Segmentation as shown in Table 1.

TABLE 1. Evaluation of segmentations methods.

Datasets	Computational Time		Segmentation Accuracy	
	GMM	AFCM	GMM	AFCM
UCM	160.13s	147.3s	83.2%	92.28%
RESIEC45	169.21s	145.8s	85.6%	92.17%

C. CONDITIONAL RANDOM FIELD(CRF) FOR LABELING

Next, Conditional Random Fields is applied on Adoptive Fuzzy C-Means segmented images for the labeling process. It is post-processing step whose purpose is to refine the segmentation results by incorporating contextual dependencies and spatial relationships between pixels. CRF is a probabilistic graphical model that considers both the unary and pairwise potentials in order to optimize and enhance the labeling process [61], [62], [63]. The energy function of the CRF is defined as:

$$E(y \mid x) = \sum_{i} \psi_u (y_i, x_i) + \sum_{i,j} \psi_p (y_i, y_j, x_i, x_j) \tag{6}$$

Here, y represents the label assignments for each pixel, and x is the observed image data. The unary potential ψ_u captures the compatibility of a label with the observed data at a single pixel, while the pairwise potential ψ_p models the compatibility between labels assigned to neighboring pixels. The goal is to minimize this energy function in order to achieve refined and coherent label assignments. The membership values which are obtained from AFCM are used to assess the certainty of the assigned labels. The pixels which contain high membership values are considered more certain and contribute in favor of the unary potential for the particular cluster [64]. The pairwise potential is designed to consider spatial relationships between neighboring pixels. It encourages label

FIGURE 5. Object labeling using conditional random field.

TABLE 2. Dice coefficient (DC) score AFCM over UCM dataset.

Objects	DC	Objects	DC	Objects	DC
AL	0.98	SK	0.98	IN	0.99
BH	0.97	AP	0.92	MK	0.91
CL	0.99	BG	0.93	PG	0.98
FT	0.86	DL	0.90	RY	0.93
GE	0.85	FY	0.86	SL	0.89
TT	0.88	HR	0.93	OS	0.85
BL	0.93	ML	0.94	RR	0.91
Mean Dice Coefficient = 92.28%					

consistency between adjacent pixels, ensuring that neighboring regions with similar characteristics share the same label. This spatial regularization helps to eliminate isolated, inconsistent labels and promotes smooth transitions between different regions in the final segmented image [65].

D. DICE COEFFICIENT

In order to evaluate the effectiveness of our segmentation method Dice coefficient is used. It works by measuring the similarity between the predicted segmentation and the ground truth as shown below.

$$Dice = \frac{2 \times |X \cap Y|}{|X| + |Y|} \tag{7}$$

Here, X represents the set of pixels in the predicted segmentation, Y is the set of pixels in the ground truth, and $|.|$ denotes the cardinality of a set [66], [67]. The Dice coefficient ranges from 0 to 1, where a higher value indicates better overlap and similarity between the predicted and ground truth segmentations [68], [69]. The results of dice coefficients for both datasets are shown in Table 2 and 3 and it shows that the implementation of conditional random field on adoptive fuzzy c mean segmented images have improved the accuracy rate signifenlty. The dice coefficient score of UCM dataset is 92.28% and 92.17% on RESICS 45 dataset.

E. FEATURE EXTRACTION

One of the most important step of this model is the feature extraction which play a vital role in classification of aerial images. The basic purpose of this step is to distill essential information from the raw pixel data. Aerial imagery, with its inherent complexity and variability, often contains high-dimensional data that can be computationally expensive. Feature extraction addresses this issue by transforming the data into a more compact and informative representation [70], [71]. This reduction in dimensionality facilitates more

Algorithm 1 Pseudo-Code for the Proposed Model

Input: Image I, Number of clusters K
Output: Segmented image
// Initialization
init_AFCM(membership values, cluster centroids)
// Adaptive Fuzzy C-Means (AFCM) Segmentation
// AFCM objective function
J = Σ (i=1)^c Σ (j=1)^n μ_ij^m ∗ d_ij ^2
converged = False
while not converged**:**
  for i in range(c):
    for j in range(n):
  // Update membership values
    μ_ij = [Σ(k=1)^c (d_ij/d_kj)^(2/(m-1))]^(-1)
    End for
  End for
  //Update cluster centroids
  update_centroids(cluster_centroids, membership_values)
  //Check convergence condition
  if convergence_criteria_met():
    converged = True
// Assign each pixel to the cluster with the highest membership value
for pixel in image:
  pixel_label = cluster_with_max_membership (pixel, membership_values)
  segmented_image[pixel] = pixel_label
End for
//Conditional Random Field (CRF) for Labeling
// CRF energy function
E(y|x) =Σ_iψ_u(y_i, x_i) +Σ_(i,j)ψ_p(y_i, y_j, x_i, x_j)
// Use membership values from AFCM to assess label certainty
for pixel in segmented_image:
  unary_potential = compute_unary_potential(pixel, membership_values)
End for
// Compute pairwise potentials for neighboring pixels
for pixel, neighbor in get_neighbors(segmented_image):
  pairwise_potential = compute_pairwise_potential (pixel, neighbor)
End for
// Optimize CRF energy function
optimize_CRF_energy(segmented_image, unary_potentials, pairwise_potentials)
Output: Final segmented image
return segmented_image

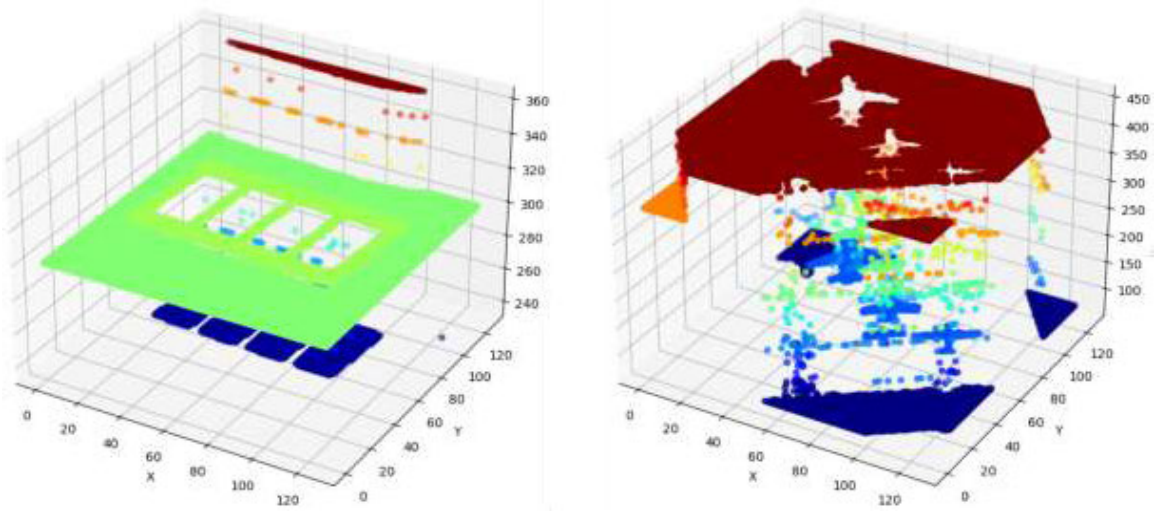


FIGURE 6. Harr wavelet transform applied on Tennis Court (left) and Air Field (right) for feature extraction.

TABLE 3. Dice coefficient (DC) score AFCM over RESISC45.

Objects	DC	Objects	DC	Objects	DC
AP	0.97	CP	0.95	CF	0.94
BR	0.97	DR	0.97	DS	0.91
CL	0.98	GC	0.91	GT	0.87
FR	0.87	IN	0.87	IS	0.90
HR	0.91	MR	0.89	MK	0.92
LK	0.96	PL	0.91	PG	0.91
MT	0.93	RF	0.88	RV	0.96
RW	0.95	SI	0.89	SP	0.94
RD	0.91	SM	0.92	ST	0.89
SB	0.89	TS	0.89	WT	0.96
TC	0.93	BS	0.93	BH	0.95
AT	0.96	SR	0.91	OP	0.89
CH	0.97	TR	0.93	RS	0.88
CM	0.90	BB	0.89	RY	0.94
FW	0.92	ID	0.94	MW	0.92
Mean Dice Coefficient = 92.17%					

efficient processing and analysis. Moreover, the extracted features enhance the discriminatory power of the data, allowing subsequent classification algorithms to differentiate between objects, terrains, or structures present in the images [72], [73]. The process also contributes to increased robustness against noise and environmental variability, which is particularly pertinent in aerial imagery where factors like lighting conditions and perspectives can introduce considerable variations. So for this purpose we used wavelet packet features, Haar wavelet, Gabor and Correlogram, as they collectively contributed to a more comprehensive analysis for subsequent classification tasks.

1) HAAR WAVELET TRANSFORM

The First feature extraction methods used in this study is Haar wavelet transform. It plays an important role in capturing variations in intensity across different scales within an image. It is important to note that Haar wavelet transform takes less time and it has the ability to identify abrupt changes in different lightning conditions. This makes it valuable in scenarios where edges and high-frequency details are of interest [74]. The Haar Wavelet Transform is applied to a two-dimensional image using a pair of low-pass *(LL)* and high-pass *(LH), (HL), (HH)* filters. The approximation coefficients *(LL)* represent the low-frequency components, while the detail coefficients *(LH), (HL),* and *(HH)* capture the high-frequency components in the horizontal, vertical, and diagonal directions, respectively [75], [76] as shown in Fig. 6 where the horizontal pixel coordinates are displayed along x-axis which depicts width of the image. Y-axis represents the vertical pixel coordinates which corresponds to the height of the image and Z-axis represents the intensity of the Haar wavelet transform.

Let I be the CRF-labeled image, and W_H represent the Haar Wavelet Transform. The transformation is performed iteratively, producing approximation and detail coefficients at each level. Mathematically, the transformation can be expressed as:

$$I^{(1)} = W_H(I) = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} I_{even} \\ I_{odd} \end{bmatrix} \tag{8}$$

Here, $I^{(1)}$ contains the approximation coefficients *LL* and detail coefficients *LH, HL, HH* of the first level. This process can be recursively applied to further decompose the approximation coefficients into subsequent levels, yielding a multi-resolution representation of the image [77]. The final outcome consists of the extracted features that capture variations in intensity at different scales and orientations.

2) WAVELET PACKET TRANSFORM (WPT)

The Wavelet Packet Transform further enhance the standard wavelet transforms as it provides a more versatile decomposition by allowing both approximation and detail coefficients to be further decomposed [78], [79], [80]. Let I be the

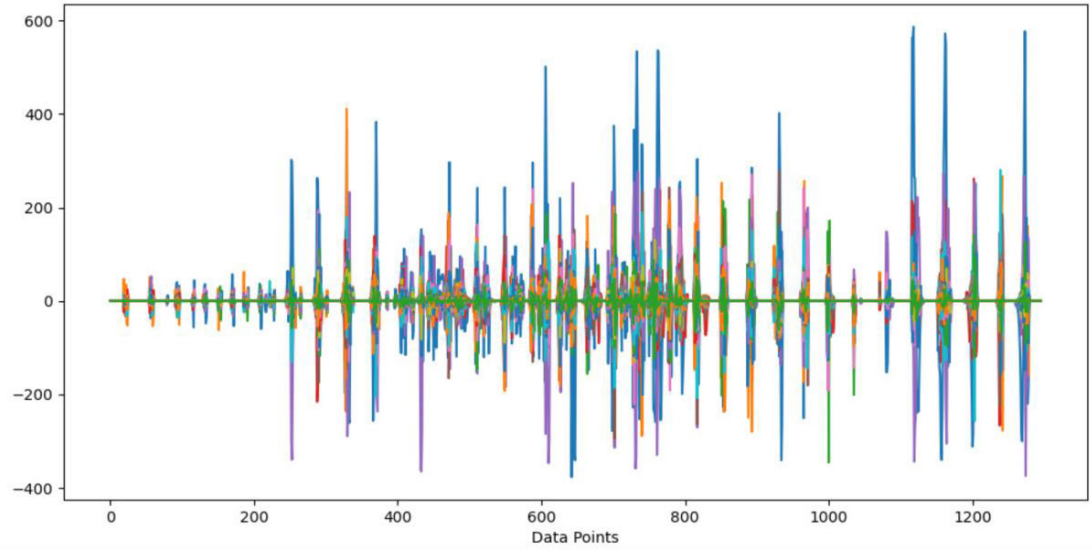


FIGURE 7. Feature extraction using wavelet transform of air field image.

CRF-labelled image, and W_{packet} represent the Wavelet Packet Transform. Unlike the Haar Wavelet Transform, which uses a fixed set of filters, the Wavelet Packet

Transform allows for more versatility in choosing the filters. The process involves recursively decomposing the approximation and detail coefficients into sub bands, resulting in a tree structure [81]. Mathematically, the Wavelet Packet Transform can be expressed as:

$$I^1 = W_{packet}\,(I) = [A^1 D^1] \tag{9}$$

Here, A^1 contains the approximation coefficients, and D^1 contains the detail coefficients at the first level. Unlike the Haar Wavelet Transform, the Wavelet Packet Transform further decomposes both approximation and detail coefficients into sub bands, providing a more detailed and comprehensive set of features [82]. The process is then recursively applied to each sub band, creating a tree structure where each node represents a different frequency sub band. The resulting coefficients at different levels constitute the extracted features, capturing variations in intensity and spatial details at multiple scales and orientations. WPT retains both high and low-frequency components and is a superior choice when inspecting finer details within an image that has been segmented. The process of separating the image into different frequency sub-bands which is represent along X-axis and the magnitude of the coefficient at each of these frequency band along Y-axis as displayed in the Fig. 7, using WPT which provides a comprehensive depiction of it which eases the job of differentiating between various regions which improves the accuracy of image classification.

3) GABOR FILTER BASED FEATURE EXTRACTION

The feature extraction process using Gabor filters involves capturing relevant information about the texture patterns present in an image. The Gabor filter is used to analyze spatial frequencies at various orientations and scales. The mathematical representation of the Gabor filter response $g(x, y)$ at a pixel location (x, y) is expressed as:

$$g\,(x, y) = e^{-\frac{X^2+\gamma^2 Y^2}{2\sigma^2}} \cos(2\pi \frac{X}{\lambda} + \phi) \tag{10}$$

In this equation, X and Y represent the rotated coordinates after applying the specified orientation, γ is the spatial aspect ratio, σ controls the spread of the Gaussian envelope, λ is the wavelength of the sinusoidal function, and ϕ is the phase offset [83]. For feature extraction, a filter bank of Gabor filters is applied across the entire image, systematically calculating responses that characterize local texture information at different orientations and scales. The responses are computed for each pixel, resulting in a set of feature vectors. These feature vectors collectively represent the texture characteristics of the entire image. The integration of these feature vectors into subsequent analyses, such as image classification or segmentation, allows for a more nuanced exploration of texture patterns within the image [84]. This feature extraction method plays a vital role for edge detections of different objects from the images making model to learn patterns within the segmented images. In the Fig. 8 the x-axis and y-axis represent the spatial coordinates within the segmented image. The x-axis corresponds to the horizontal pixel coordinates, while the y-axis corresponds to the vertical pixel coordinates. The z-axis, on the other hand, represents the probability density values computed from the multivariate normal distribution (Gaussian distribution) fitted to the pixel coordinates of each segmented object within the image.

4) CORRELOGRAM FEATURES

Correlogram features are employed for texture analysis in image processing, providing insights into the spatial

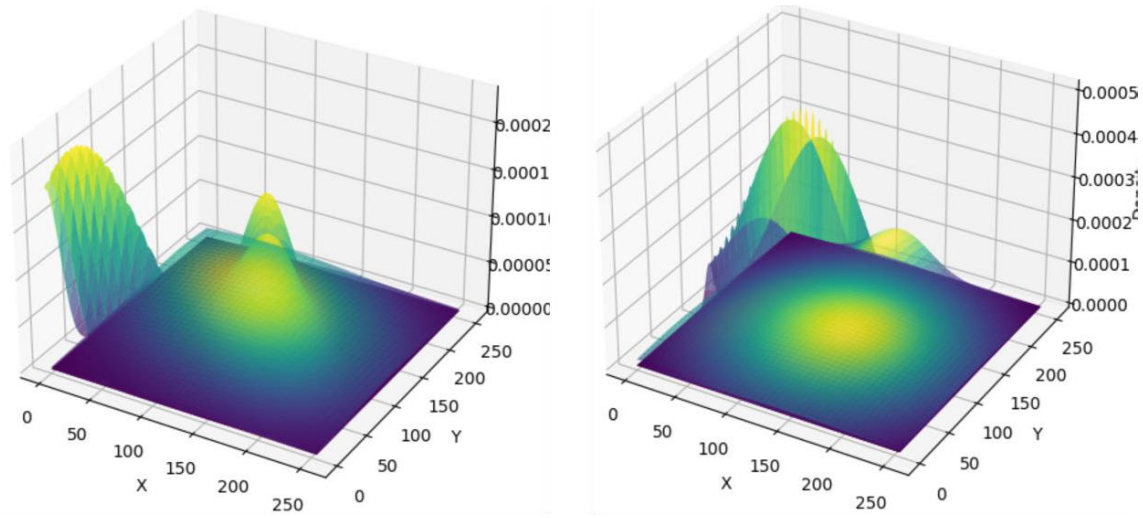


FIGURE 8. Feature extraction using gabor on left is tennis court and on right is air field.

relationships between pixel intensities at varying distances and angles. It quantifies the probability of having a pair of pixels with similar intensity values at a given distance and angle in an image. It is important in order to characterizing the texture and structural patterns present in aerial images.

Let I be the CRF-labelled aerial image, and $P(d,\theta)$ denote the correlogram, where d represents the distance between pixel pairs, and θ signifies the angle. The computation of the correlogram involves counting the number of pixel pairs with similar intensity values within specified distance and angle bins [85], [86]. Mathematically, the correlogram can be expressed as:

$$\mathrm{P}\,(\mathrm{d},\theta) = \frac{1}{N_d(\theta)} \sum_{i=1}^{N_d(\theta)} \delta\left(I\left(x_i\right) - I\left(x_i + d\theta\right)\right) \qquad (11)$$

Here $N_d(\theta)$ is the number of pixel pairs separated by distance d and angle θ. $\mathrm{I(x_i)}$ and $\mathrm{I(xi + d\theta)}$ are the intensity values of the pixel pairs at positions xi and $\mathrm{xi + d\theta)}$, respectively and $\delta(.)$ is the Kronecker delta function, yielding 1 if the intensity values are similar and 0 otherwise.

Correlogram features are typically computed for various distance and angle combinations, creating a multi-dimensional feature vector that characterizes the spatial relationships in the image. With the assistance of correlogram, the classifier can build a better capacity of distinguishing the similar texture regions yet depicting non-corresponding color distribution, hence enhancing the robustness of the classification model.

It is important to note that each feature extraction method plays a unique role in the classification process by capturing distinct aspects of segmented images. The Wavelet Packet Transform (WPT) offers a detailed frequency breakdown, enhancing the representation of intricate textures. The Haar Wavelet Transform efficiently detects edges and sudden intensity changes crucial for boundary delineation. Gabor filters excel in extracting spatial frequency characteristics across various orientations and scales, aiding in texture and edge identification. Correlograms capture spatial color correlations, providing statistical insights into color patterns and distributions. These techniques were chosen for their proven efficacy in capturing diverse image features, creating a comprehensive and discriminative feature set that significantly boosts classification accuracy.

F. FEATURE FUSION AND OPTIMIZATION

Feature fusion is a crucial step in leveraging complementary information from diverse feature extraction methods to enhance the overall effectiveness of a system. In this context, concatenation serves as a straightforward yet powerful approach to fuse features extracted from Gabor filters, Wavelet Packet Transform, Haar Wavelet Transform, and Correlogram features. Concatenation combines the distinct characteristics of each feature set into a unified feature vector, providing a comprehensive representation of the image content [87], [88]. Mathematically, concatenated feature vector Fused is given by:

$$Fused = [F_{Gabor}, F_{Wavelet}, F_{Haar}, F_{Correlogram}] \qquad (12)$$

This concatenated feature vector effectively captures the combined information from multiple feature extraction methods, paving the way for a more comprehensive analysis.

To further optimize the performance of the fused features, Genetic Algorithms (GAs) are employed as a feature selection and dimensionality reduction technique [89], [90]. GA has the capability to explore large search space effectively and they have the capability to find the global optima where there is complex and high dimensional feature space. Moreover, it has the advantage of handling the nonlinear relationships in the data which some of the other optimization methods cannot. Due to their ability to evaluate different solutions simultaneously, the convergence becomes faster. These advantage of GA makes it an optimal choice for our

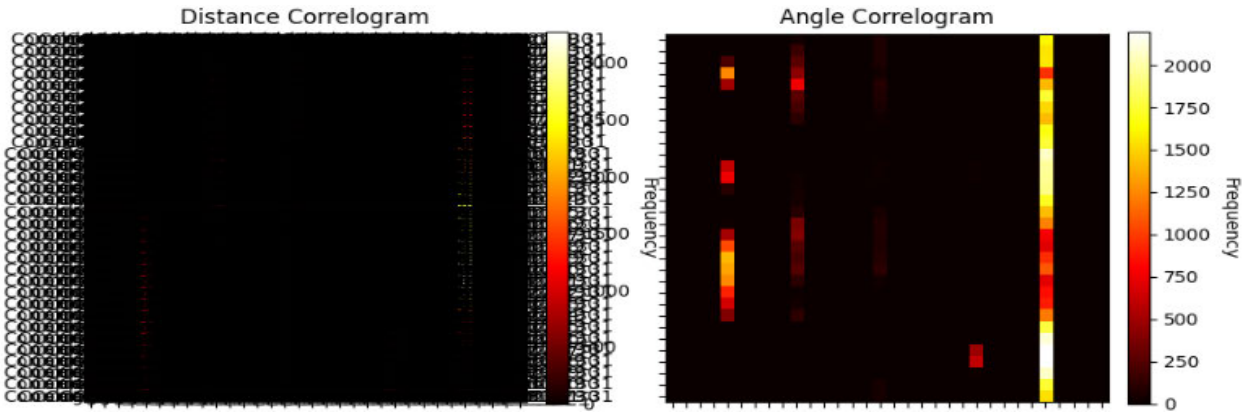


FIGURE 9. Feature extraction using correlogram from tennis court showing the distance and angle correlogram.

work to be used for optimization purpose. GA's iteratively evolve a population of potential solutions based on natural selection principles, seeking an optimal subset of features that maximizes a defined objective function. In this case, the objective function could be related to the performance of a classification or analysis task. The optimization process involves encoding the concatenated feature vector into a chromosome, and the genetic algorithm performs operations such as selection, crossover, and mutation to iteratively improve the feature subset. The output is an optimized subset of features that exhibits enhanced [91], [92], [93].

G. CLASSIFICATION: DEEP BELIEF NETWORK

For the classification process we have employed Deep belief network. It's noteworthy we utilized GA to optimize the features which we have already extracted and these optimized features are feed as input to DBN which ensures that supplied to the DBN are fine-tuned for optimal performance. Deep belief Network is a problastic model have multiple hidden interconnected layer and based on artificial neural network model [94]. These interconnected layers contribute to the refinement of unit weights through an iterative process [95], [96], [97]. Theses weights can be adjusted using the following equation.

$$Wi, j\,(t+1) = Wi, j\,(t) + n\log(P(v)) \tag{13}$$

And the probability p(v) can be calculated as

$$P\,(v) = \frac{1}{Z}\sum\nolimits_{h} e^{-E(v,h)^{-}} \tag{14}$$

The weight adjustment equation $W\,i,j(t+1)$ captures the incremental refinement of weights based on the logarithm of the probability of the visible vector. The probability p(v) is fundamental to the DBN's learning process. It sums over the exponential terms of the energy function (v, h), which encapsulates the interactions between the visible and hidden layers. By exponentiation and summing these terms, the equation normalizes the probabilities across all possible configurations

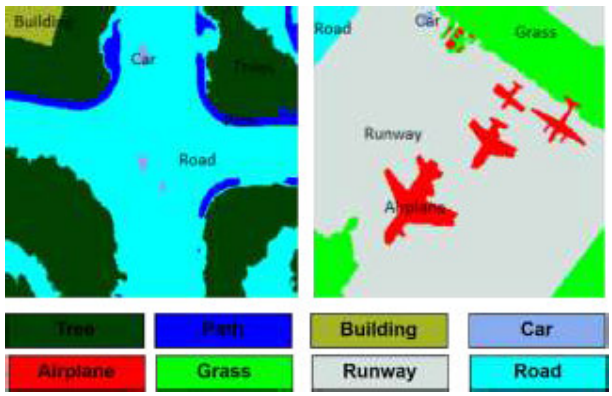


FIGURE 10. Multi object classification using our proposed model.

of the hidden layers. More informally, equation 14 specifies the probability distribution over the visible vector based on the current state of the hidden layers. This probabilistic property enables the DBN to learn intricate probabilistic relations and patterns from the input data thus enabling accurate classification.

The DBN's architecture contains visible and output layers, along with three hidden layers. Genetic Algorithm is used to determine optimal units for each hidden layer and total epochs in network [2], [98], [99]. The output layer is responsible for generating class probabilities based on input values which are object categories, IOU scores, and extracted features. The training process of our DBN includes several stages. Layer-wise unsupervised pre-training serves to initialize the network weights, for which each layer of the Restricted Boltzmann Machine is trained using Contrastive Divergence. These methods maximize correlation by enhancing the network's ability to learn meaningful patterns from input data. After pre-training, back-propagation is used in tightening the entire network, where weights are regulated to reduce classification error in this supervised training. Cognitive parameter weights are adjusted during pre-training and fine-tuning phases by stochastic gradient

FIGURE 11. Few of the example images from RESISC45 and UCM datasets.

descent algorithm. However, for RBM layers we use the CD-k algorithm which is often set at k = 1. The stochastic operations introduce randomness which allows the network to explore different configurations and converge to a robust solution. During training, the Softmax and modified Sigmoid functions are used for multi-object classification. Unlike traditional Softmax, the modified Sigmoid treats each neuron as an independent binary classifier, capturing the probability of individual object classes simultaneously. It continues until one of three conditions is met: achieving the maximum epoch of 200, approaching the minimum gradient, or meeting the mean squared error benchmark. The GA, optimizing hidden units, utilizes a fitness function based on misclassification rate, total training time before backpropagation, and time taken for fine-tuning during backpropagation. Optimal convergence of the fitness function is declared when both training time and error rates are minimized. After 25 generations, evolution ceases, and the selected optimized neurons in the three hidden layers (800, 450, and 1580) and epochs generated by the genetic algorithm (120, 100, and 136) are employed for object classification. The DBN classifier generates probability values for input feature vectors, facilitating training against specific datasets with corresponding class numbers [100]. There are several distinct advantages of using Deep Belief Networks (DBNs) for classification tasks, particularly when input values are in the form of pre-extracted features. Here, the DBN is able to learn the complexity within the features without needing to extract features on its own by feeding in pre-processed features as the input for the network. This reduces computational complexity in certain domains and also the time taken to train the classifiers thus making the classification easier. Another advantage of DBNs is their ability to learn a hierarchical representation, which allows the discrimination of both the high-level semantic features and the low-level details that may be present in the features. Finally, DBNs are flexible, and the method is very efficient in classification no matter the type of data being fed into the network. The object wise classification of the object is shown in Fig. 10.

IV. EXPERIMENTATION AND RESULTS

The experiments were carried out on a system equipped with an Intel Core i3 processor running at 1.7 GHz, operating on Windows 10. For the implementation and analysis, we used Python. The datasets are divided into the ration of 70:30 for the training and testing respectively. The proposed model produces remarkable results on UCM and RESISC45 benchmark datasets. Details of the dataset and evaluation is discussed in the section given below.

A. DATASETS

1) UCM DATASET

The UCM [101] dataset is a high-resolution remote sensing dataset for multi object classification. The data set contains aerial images captured covering urban, rural, and other landscapes. It contains total of 21 land-use classes, including residential and agricultural areas. Each class contains 100 images while the dimension of the images is 256 X 256. Moreover, the number of object in each image are not consistent making it a good choice for multi object classification.

2) RESISC45 DATASET

The RESISC45 dataset [102] is a collection of 31,500 remote sensing images. These images are divided into 45 distinct classes where the number of images in each class are 700. Also the number of objects are different in each image. The dataset is best suited for multi object classification in remote sensing images. It contains diverse classes like airports

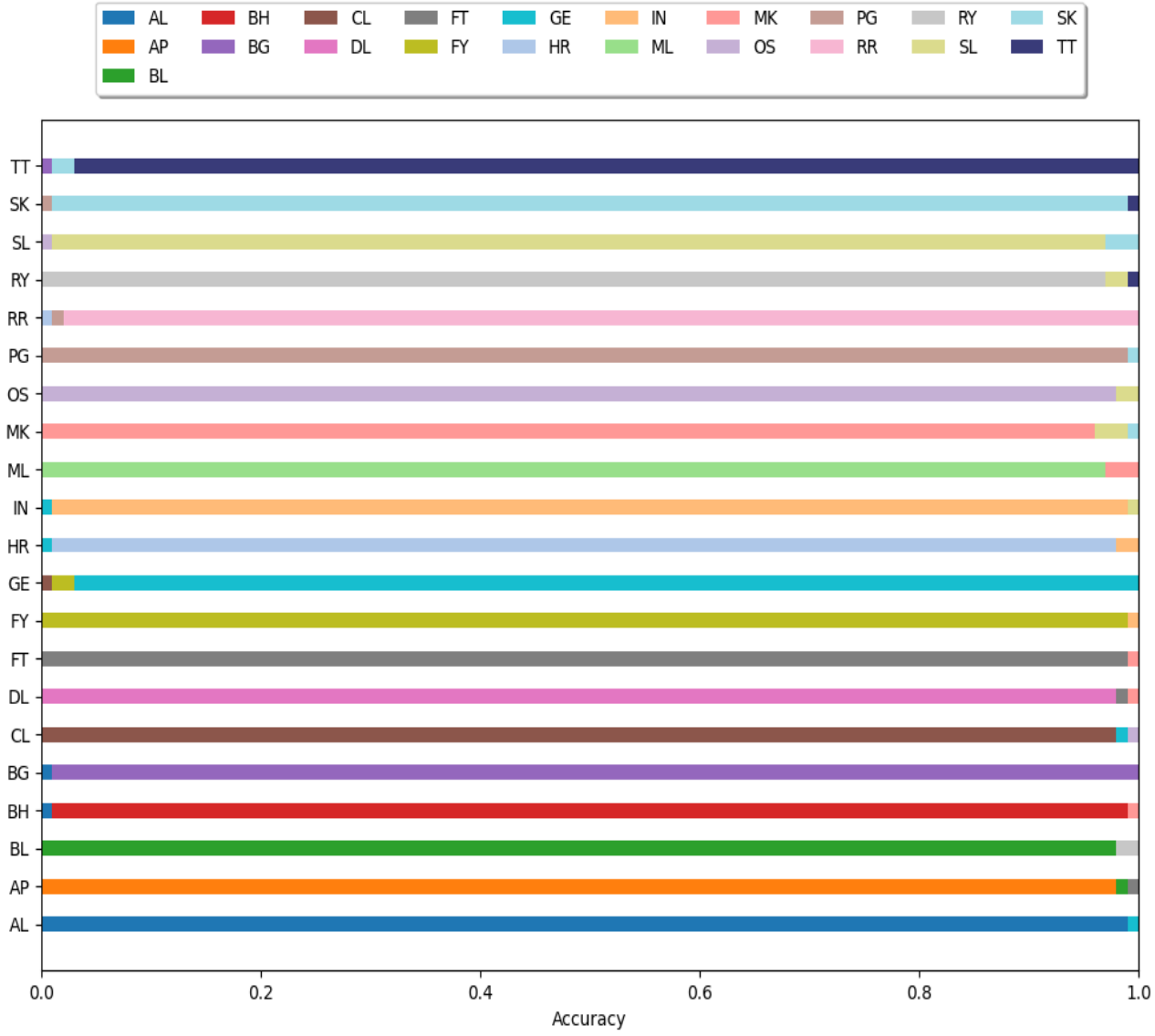


FIGURE 12. The recognition accuracy of over UCM dataset. AL = agricultural; AP = airplane; BL = baseball diamond; BH = beach; BG = building; CL = chaparral; DL = dense residential; FT = forest; FY = freeway; GE = golf course; HR = harbor; IN = intersection; ML = medium residential; MK = mobile home park; OS = overpass; PG = parking; RR = river; RY = runway; SL = sparse residential; SK = storage tank; TT = tennis court.

playing area, industrial area and urban areas. RESISC45 has become a benchmark dataset for research in object classification algorithms.

B. EVALUATION AND DISCUSSION

In this section, we present the achieved recognition accuracies in form of confusion matrices across UCM, and RESISC45 datasets. Results, visually represented in Fig. 12, and 13, showcase the system's robust performance. Specifically, an outstanding 96.25% average accuracy in image classification was observed over the UCM dataset. The agriculture, parking, freeway, forest, intersections and building classes exhibits the accuracy of 99%. While sparse residential and medium residential strugglers with 96% accuracy comparing with other classes better accuracy due to much of the similarity. Furthermore, RESISC45 dataset demonstrated an average accuracy of 97.13%. Our result shows that the model performed well comparing to most of the state of the art methods. One of the major factor is that our segmentation method AFCM produced good results which are improved more in post processing step using CRF for labeling which helped to get better feature extraction. Moreover, our feature extraction methods used in this paper proved to be very effective as they contributed in achieving the high accuracy in multi object classification.

In this extensive analysis, we conducted a thorough evaluation of precision (Pn.), recall (Rc.), and F1 scores obtained from comparing RNN and DBN models on two datasets RESISC45 and UCM. Table 4 on UCM dataset represents the mean precision, recall, and F1 scores for RNN are

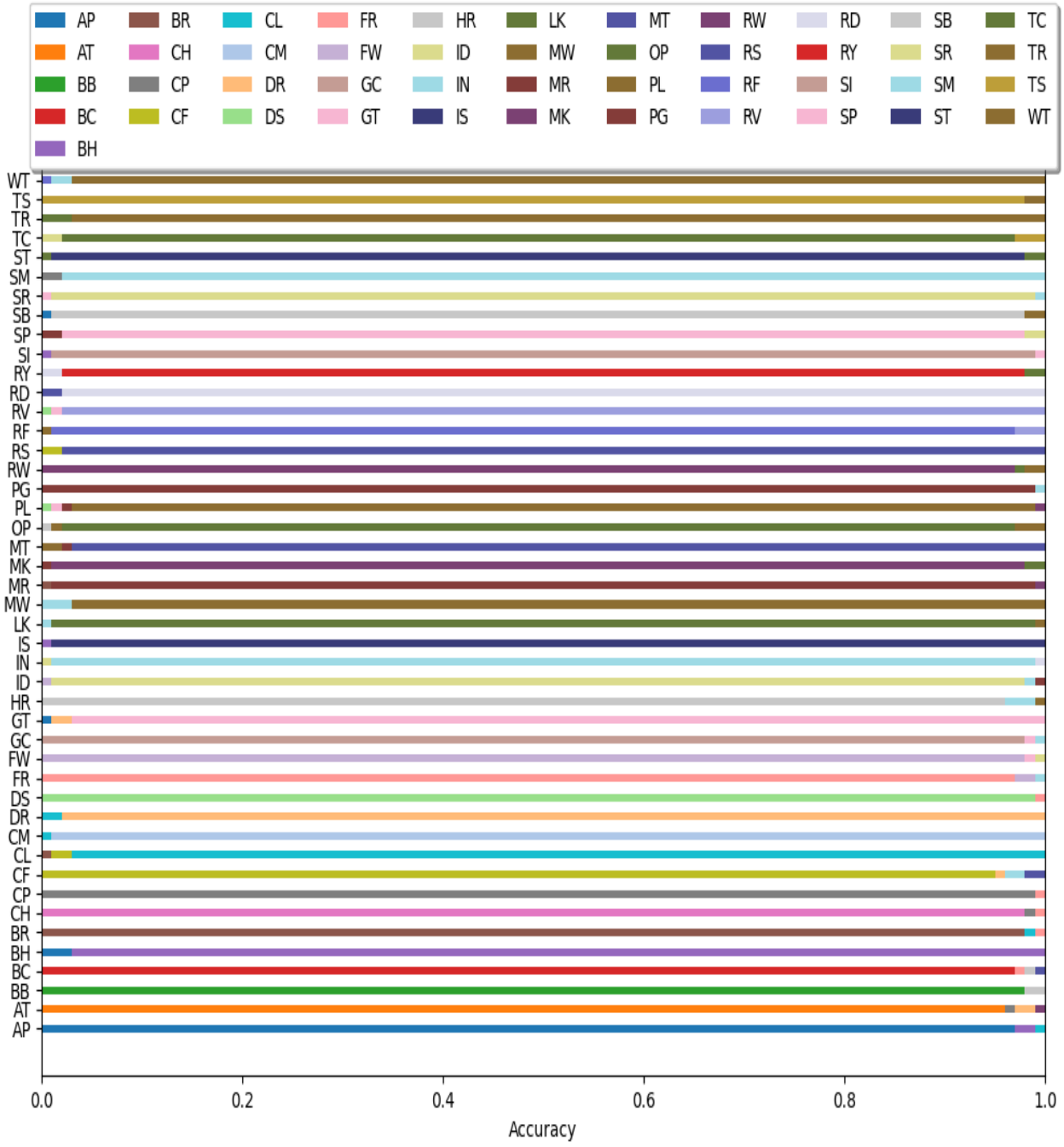


FIGURE 13. The recognition accuracy of model over RESISC45 dataset. AP = airplane; AT = airport; BB = baseball diamond; BC = basketball court; BH = beach; BR = bridge; CH = church; CP = chaparral; CF = circular formland; CL = cloud; CM = commercial area; DR = dense residential; DS = desert; FR = forest; FW = freeway; GC = golf course; GT = ground track field; HR = harbor; ID = island; IN = industrial area; IS = intersection; LK = lake; MW = meadow; MR = medium residential; MK = mobile home park; MT = mountain; OP = overpass; PL = palace; PG = parking lot; RW = railway; RS = railway station; RF = rectangular formland; RV = river; RD = roundabout; RW = runway; SI = sea ice; SP = ship; SB = snow berg; SR = sparse residential; SM = stadium; ST = storage tank; TC = tennis court; TR = terrace; TS = thermal power station; WT = wetland.

0.912, 0.891, and 0.945, respectively. In comparison, DBN achieves mean scores of 0.921, 0.905, and 0.915 for precision, recall, and F1. These results show the robust and consistent performance of both models across the diverse set of object categories present in the UCM dataset. Similarly, when examining the RESISC45 dataset as represented in Table 5, RNN demonstrates mean precision, recall, and F1 scores of 0.852, 0.811, and 0.885 while, DBN exhibits mean scores of 0.921, 0.905, and 0.915 for precision, recall, and F1. The overall mean values emphasize the models' competence in

TABLE 4. Precision (Pn.), Recall (RC.) and F1 Score (F1 S) Object categorization (OC) among RNN and DBN UCM dataset.

Classes	RNN			DBN		
	Pn.	Rc.	F1 S	Pn.	Rc.	F1 S
AG	0.755	0.788	0.755	0.799	0.899	0.817
AP	0.711	0.744	0.875	0.815	0.875	0.844
BB	0.783	0.711	0.746	0.841	0.819	0.747
BH	0.792	0.658	0.737	0.844	0.889	0.844
BD	0.701	0.725	0.758	0.889	0.839	0.889
CH	0.745	0.715	0.875	0.872	0.921	0.872
DN	0.799	0.791	0.795	0.886	0.938	0.886
FR	0.783	0.711	0.746	0.985	0.954	0.985
FW	0.771	0.792	0.781	0.901	0.859	0.901
GC	0.730	0.717	0.961	0.883	0.965	0.883
HR	0.755	0.788	0.755	0.879	0.851	0.879
IN	0.711	0.744	0.875	0.986	0.937	0.986
MR	0.783	0.711	0.746	0.967	0.809	0.967
MH	0.792	0.658	0.737	0.845	0.856	0.850
OP	0.701	0.725	0.758	0.879	0.851	0.864
PN	0.874	0.845	0.859	0.986	0.937	0.960
RV	0.869	0.829	0.903	0.967	0.809	0.880
RW	0.872	0.851	0.895	0.844	0.855	0.850
SP	0.886	0.918	0.911	0.899	0.839	0.867
SN	0.965	0.934	0.969	0.872	0.875	0.873
TC	0.901	0.859	0.937	0.886	0.913	0.899
Mean	**0.912**	**0.891**	**0.945**	**0.921**	**0.905**	**0.915**

dealing with the complexities of object categorization within this dataset. The consistency in performance across diverse categories underlines the models' adaptability and effectiveness in handling the intricacies present in both the UCM and RESISC45 datasets.

Table 5. displays the comparison the mean accuracy of our proposed model with state of the art methods. As it can be seen that our model performed well on both the datasets comparatively to most of the methods. Although the proposed model by OSCS [31] and Resent 18 [103] based model performed better on UCM dataset than our model but our model performed better than their model on REISSC 45 dataset which showcases that our model is robust and can perform equally well on different datasets and can identify the complex patterns from the given datasets. Moreover, the results depict that our feature set is very effective for further classification of the objects from images. One important factor that took major part in better results is our segmentation methods in which we used adoptive fuzzy c mean and then with the help of post processing step using CRF for the labeling have boosted the overall performance of our proposed model.

C. LIMITATIONS

Although, our model performed outstanding results on both the datasets and produced State of the Art results on RESIC45 datasets but it also exhibits few limitations. The model's

TABLE 5. Results for OC among RNN and DBN RESISC45 dataset.

classes	RNN			DBN		
	Pn.	Rc.	F1 S	Pn.	Rc.	F1 S
AP	0.843	0.752	0.914	0.864	0.872	0.906
BR	0.898	0.743	0.925	0.842	1.011	0.947
CL	0.911	0.888	0.885	0.980	0.925	0.948
FR	0.950	0.757	0.838	0.856	0.836	0.930
HR	0.774	0.831	0.916	1.032	0.928	0.944
LK	0.952	0.854	0.907	1.008	0.952	0.868
MT	0.958	0.772	0.851	0.961	0.901	0.900
RW	0.782	0.853	0.863	0.929	0.955	0.923
RD	0.968	0.771	0.910	0.897	0.990	0.957
SB	0.769	0.804	0.909	0.826	0.866	0.963
TC	0.874	0.888	0.894	0.876	0.953	0.897
AT	0.757	0.756	0.839	0.850	0.828	0.922
CH	0.896	0.866	0.845	0.855	1.017	0.913
CM	0.808	0.894	0.898	0.830	0.872	0.900
FW	0.764	0.746	0.860	0.859	0.961	0.880
ID	0.925	0.897	0.841	0.920	0.918	0.878
MW	0.754	0.894	0.864	0.928	1.012	0.937
OP	0.801	0.864	0.918	1.017	0.804	0.907
RS	0.800	0.845	0.910	0.917	0.823	0.920
RY	0.810	0.744	0.832	0.980	0.902	0.885
SR	0.844	0.830	0.904	0.922	0.961	0.857
TR	0.797	0.759	0.923	1.001	0.971	0.862
BB	0.897	0.719	0.851	0.986	0.809	0.944
CP	0.879	0.879	0.889	1.029	0.938	0.876
DR	0.935	0.814	0.838	0.971	0.931	0.879
GC	0.773	0.900	0.860	0.825	0.967	0.950
IN	0.899	0.766	0.924	0.897	0.865	0.927
MR	0.923	0.769	0.898	0.893	0.909	0.888
PL	0.954	0.719	0.900	0.981	0.986	0.857
RF	0.817	0.896	0.888	0.827	0.994	0.962
SI	0.812	0.776	0.897	0.863	0.871	0.943
SM	0.792	0.755	0.917	0.966	0.909	0.942
TS	0.896	0.818	0.895	0.993	0.815	0.936
BC	0.918	0.845	0.900	0.838	0.829	0.880
CF	0.882	0.746	0.892	0.873	0.920	0.932
DS	0.866	0.901	0.896	0.835	0.871	0.928
GT	0.934	0.883	0.841	1.024	0.874	0.876
IS	0.771	0.855	0.866	0.831	0.985	0.953
MK	0.764	0.827	0.910	0.906	0.865	0.897
PG	0.760	0.734	0.888	1.031	0.832	0.889
RV	0.858	0.873	0.866	0.984	0.824	0.878
SP	0.859	0.773	0.903	0.842	0.809	0.962
ST	0.783	0.768	0.915	0.951	0.877	0.957
WT	0.859	0.713	0.901	0.983	0.945	0.940
BH	0.874	0.757	0.846	0.965	0.842	0.931
Mean	**0.852**	**0.811**	**0.885**	**0.921**	**0.905**	**0.915**

performance got degraded in the classes which are visually very much similar like the sparse residential and medium residential in UCM dataset, or industrial and commercial areas in RESISC45. Similarly, classes with lot of texture like beach and desert also got misclassified occasionally, and has the room for further improvement. Some of the misclassifications are result of the scale invariant objects as same objects have

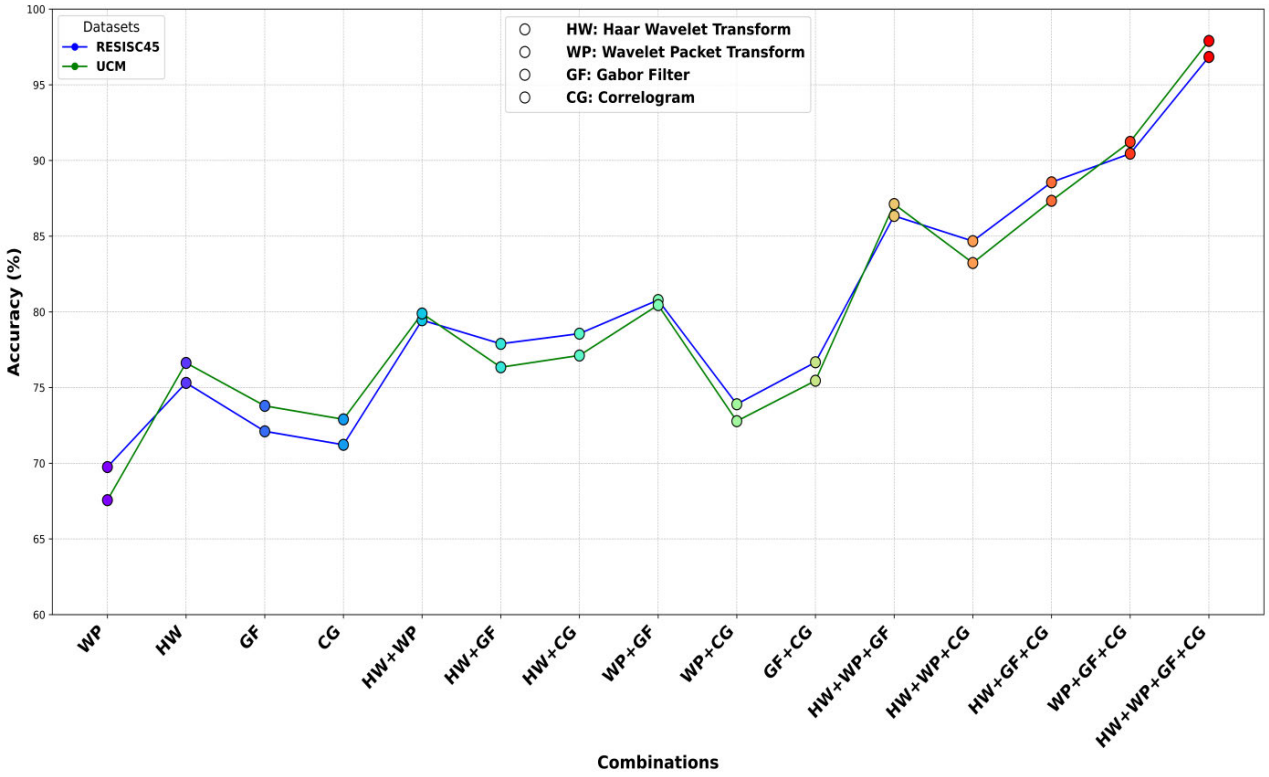


FIGURE 14. Accuracy of feature extraction methods and their combinations on our proposed model on RESISC45 and UCM datasets.

TABLE 6. Comparison of our model with SOTA methods.

Authors	Mean Accuracies %	
	RESISC45	UCM
Self-attention [104]	86.91	86.79
Feature fusion + ELM [105]	96.97	84.00
FSCNet with CRIM[106]	94.76	99.60
SSGA-E[107]	88.6	94.52
D-CNN with GoogLeNet[108]	90.49	97.07
RS-CLIP [109]	85.76	95.94
VHR RS[110]	93.04	98.61
Siamse ResNet50[111]	95.95	94.39
OCSC[31]	96.57	98.75
ResNet18 + LA + KL[103]	95.26	99.21
Proposed	**96.84**	**97.90**

different size, distance and angles in same or different scenes. One of the key improvement or enhancement which in future we will work on is the integration of contextual information into the classification because few of the misclassification are due to the fact that the distinctive features are less prominent like gold course are misclassified as park.

D. ABLATION STUDY

To analyze the performance of our system based on feature extraction techniques, we implemented four feature extractors such as Haar Wavelet Transform, Wavelet Packet Transform, Gabor Filter, and Correlogram. The results of these feature extractors for object classification tasks are depicted in Fig. 14 across two datasets to include RESISC45 and UCM.

For the RESISC45 dataset, the individual techniques achieved accuracies ranging from 69.75% (HW) to 75.31% (WP). The combination of WP and GF features (WP+GF) yielded an accuracy of 80.78%, outperforming other two-feature combinations. Incorporating all four techniques (HW+WP+GF+CG) resulted in the highest accuracy of 96.84%, indicating the complementary nature of the diverse feature representations.

On the UCM dataset, a similar trend was observed, with individual technique accuracies ranging from 67.56% (HW) to 76.63% (WP). The WP+GF combination achieved 80.45% accuracy, while the four-technique combination (HW+WP+GF+CG) attained the highest accuracy of 97.90%. Among the three-feature combinations, WP+GF+CG consistently performed better than other combinations, with accuracies of 90.45% and 91.23% on RESISC45 and UCM, respectively. The addition of HW features to this combination marginally improved the accuracy, suggesting complementary information from HW features.

In particular, the application of the WP technique aimed at achieving higher accuracy and proved to be significantly higher compared to other individual techniques in both datasets while determining the relevant information for the classification. Additionally, the CG technique's performance was not very good when used individually, yet it was helpful when fused with other techniques, which indicates its information-complementing capacity. The obtained performance trends in both datasets sustain the effectiveness of the presented feature extraction and combination techniques. The removal of the feature fusion component reveals that the experimental setup benefits from the use of multiple feature extraction techniques, as opposed to solely relying on any one of them in isolation.

V. CONCLUSION

In this study, we presented a novel methodological approach for multi-object categorization in remote sensing. This method was tested using UCM and RESISC45 datasets. When it comes to segmentation, the models use Adoptive Fuzzy C-means, followed by CRF for labeling, and finally multi-feature technique is used for extraction The optimization process is performed with the help genetic algorithms, and multi- object classification deep belief network is used. The effectiveness of our algorithm is evident from the fact that it provides an impressive accuracy of 97.90% on UCM and 96.84% on RESISC45. As part of our ongoing research, we will explore ways to improve optimization for multiple data types and address real-time processing issues.

REFERENCES

[1] C. Galleguillos and S. Belongie, "Context based object categorization: A critical survey," *Comput. Vis. Image Understand.*, vol. 114, no. 6, pp. 712–722, Jun. 2010.

[2] M. W. Ahmed, N. A. Almujally, A. Alazeb, A. Algarni, and J. Park, "Enhanced object detection and classification via multi-method fusion," *Comput., Mater. Continua*, vol. 79, no. 2, pp. 3315–3331, 2024.

[3] Y. Xu, M. Wei, and M. M. Kamruzzaman, "Inter/intra-category discriminative features for aerial image classification: A quality-aware selection model," *Future Gener. Comput. Syst.*, vol. 119, pp. 77–83, Jun. 2021.

[4] W. Zou, Y. Sun, Y. Zhou, Q. Lu, Y. Nie, T. Sun, and L. Peng, "Limited sensing and deep data mining: A new exploration of developing city-wide parking guidance systems," *IEEE Intell. Transp. Syst. Mag.*, vol. 14, no. 1, pp. 198–215, Jan. 2022, doi: 10.1109/MITS.2020.2970185.

[5] C. Zheng, Y. An, Z. Wang, H. Wu, X. Qin, B. Eynard, and Y. Zhang, "Hybrid offline programming method for robotic welding systems," *Robot. Comput.-Integr. Manuf.*, vol. 73, Feb. 2022, Art. no. 102238, doi: 10.1016/j.rcim.2021.102238.

[6] C. Zheng, Y. An, Z. Wang, X. Qin, B. Eynard, M. Bricogne, J. Le Duigou, and Y. Zhang, "Knowledge-based engineering approach for defining robotic manufacturing system architectures," *Int. J. Prod. Res.*, vol. 61, no. 5, pp. 1436–1454, Mar. 2023, doi: 10.1080/00207543.2022.2037025.

[7] Y. Hua, L. Mou, and X. X. Zhu, "Relation network for multilabel aerial image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 4558–4572, Jul. 2020.

[8] C. Kyrkou and T. Theocharides, "EmergencyNet: Efficient aerial image classification for drone-based emergency monitoring using atrous convolutional feature fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 1687–1699, 2020.

[9] M. Alkhelaiwi, W. Boulila, J. Ahmad, A. Koubaa, and M. Driss, "An efficient approach based on privacy-preserving deep learning for satellite image classification," *Remote Sens.*, vol. 13, no. 11, p. 2221, Jun. 2021.

[10] M. A. Haq, G. Rahaman, P. Baral, and A. Ghosh, "Deep learning based supervised image classification using UAV images for forest areas classification," *J. Indian Soc. Remote Sens.*, vol. 49, no. 3, pp. 601–606, Mar. 2021.

[11] M. Abbas, S. Saleem, F. Subhan, and A. Bais, "Feature points-based image registration between satellite imagery and aerial images of agricultural land," *TURKISH J. Electr. Eng. Comput. Sci.*, vol. 28, no. 3, pp. 1458–1473, May 2020.

[12] G. Sun, L. Sheng, L. Luo, and H. Yu, "Game theoretic approach for multipriority data transmission in 5G vehicular networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 12, pp. 24672–24685, Dec. 2022, doi: 10.1109/TITS.2022.3198046.

[13] G. Sun, Y. Zhang, H. Yu, X. Du, and M. Guizani, "Intersection fog-based distributed routing for V2V communication in urban vehicular ad hoc networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 6, pp. 2409–2426, Jun. 2020, doi: 10.1109/TITS.2019.2918255.

[14] G. Sun, L. Song, H. Yu, V. Chang, X. Du, and M. Guizani, "V2V routing in a VANET based on the autoregressive integrated moving average model," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 908–922, Jan. 2019, doi: 10.1109/TVT.2018.2884525.

[15] Z. Zhang, Y. Xu, J. Song, Q. Zhou, J. Rasol, and L. Ma, "Planet craters detection based on unsupervised domain adaptation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 5, pp. 7140–7152, Oct. 2023, doi: 10.1109/TAES.2023.3285512.

[16] B. Petrovska, E. Zdravevski, P. Lameski, R. Corizzo, I. Štajduhar, and J. Lerga, "Deep learning for feature extraction in remote sensing: A case-study of aerial scene classification," *Sensors*, vol. 20, no. 14, p. 3906, Jul. 2020.

[17] G. Sun, Y. Zhang, D. Liao, H. Yu, X. Du, and M. Guizani, "Bus-trajectory-based street-centric routing for message delivery in urban vehicular ad hoc networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, pp. 7550–7563, Aug. 2018, doi: 10.1109/TVT.2018.2828651.

[18] B. Rasti, D. Hong, R. Hang, P. Ghamisi, X. Kang, J. Chanussot, and J. A. Benediktsson, "Feature extraction for hyperspectral imagery: The evolution from shallow to deep: Overview and toolbox," *IEEE Geosci. Remote Sens. Mag.*, vol. 8, no. 4, pp. 60–88, Dec. 2020.

[19] A. Sungheetha and R. Sharma, "Classification of remote sensing image scenes using double feature extraction hybrid deep learning approach," *J. Inf. Technol. Digit. World*, vol. 3, no. 2, pp. 133–149, Jul. 2021.

[20] Y. Yao, B. Zhao, J. Zhao, F. Shu, Y. Wu, and X. Cheng, "Anti-jamming technique for IRS aided JRC system in mobile vehicular networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 9, pp. 12550–12560, Sep. 2024, doi: 10.1109/TITS.2024.3384038.

[21] W.-L. Liu, J. Zhong, P. Liang, J. Guo, H. Zhao, and J. Zhang, "Towards explainable traffic signal control for urban networks through genetic programming," *Swarm Evol. Comput.*, vol. 88, Jul. 2024, Art. no. 101588, doi: 10.1016/j.swevo.2024.101588.

[22] L. Yin, L. Wang, L. Ge, J. Tian, Z. Yin, M. Liu, and W. Zheng, "Study on the thermospheric density distribution pattern during geomagnetic activity," *Appl. Sci.*, vol. 13, no. 9, p. 5564, Apr. 2023, doi: 10.3390/app13095564.

[23] Q. Wu, D. Feng, C. Cao, X. Zeng, Z. Feng, J. Wu, and Z. Huang, "Improved mask R-CNN for aircraft detection in remote sensing images," *Sensors*, vol. 21, no. 8, p. 2618, Apr. 2021.

[24] A. Naseer and A. Jalal, "Integrating semantic segmentation and object detection for multi-object labeling in aerial images," in *Proc. ICACS*, 2024, pp. 1–9.

[25] X. Li, F. Xu, H. Gao, F. Liu, and X. Lyu, "A frequency domain feature-guided network for semantic segmentation of remote sensing images," *IEEE Signal Process. Lett.*, vol. 31, pp. 1369–1373, 2024.

[26] N. Zhou, J. Hong, W. Cui, S. Wu, and Z. Zhang, "A multiscale attention segment network-based semantic segmentation model for landslide remote sensing images," *Remote Sens.*, vol. 16, no. 10, p. 1712, May 2024.

[27] X. Li, F. Xu, F. Liu, Y. Tong, X. Lyu, and J. Zhou, "Semantic segmentation of remote sensing images by interactive representation refinement and geometric prior-guided inference," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5400318.

[28] K. Chen, C. Liu, H. Chen, H. Zhang, W. Li, Z. Zou, and Z. Shi, "RSPrompter: Learning to prompt for remote sensing instance segmentation based on visual foundation model," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 4701117.

[29] S. N. Sulaiman and N. A. Mat Isa, "Adaptive fuzzy-K-means clustering algorithm for image segmentation," *IEEE Trans. Consum. Electron.*, vol. 56, no. 4, pp. 2661–2668, Nov. 2010.

[30] R. Jia, Y. Cao, H. Tang, F. Fang, C. Cao, and S. Wang, "Neural extractive summarization with hierarchical attentive heterogeneous graph network," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2020, pp. 3622–3631.

[31] Y. Y. Ghadi, A. A. Rafique, T. A. Shloul, S. A. Alsuhibany, A. Jalal, and J. Park, "Robust object categorization and scene classification over remote sensing images via features fusion and fully convolutional network," *Remote Sens.*, vol. 14, no. 7, p. 1550, Mar. 2022.

[32] K. Zhang, "Exploring hyperspectral and very high spatial resolution imagery in vegetation characterization," Dept. Earth Space Sci., York Univ., Toronto, ON, Canada, Rep., 2014.

[33] X. Li, F. Xu, L. Li, N. Xu, F. Liu, C. Yuan, Z. Chen, and X. Lyu, "AAFormer: Attention-attended transformer for semantic segmentation of remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 21, pp. 1–5, 2024.

[34] M. Rahaman, M. M. Hillas, J. Tuba, J. F. Ruma, N. Ahmed, and R. M. Rahman, "Effects of label noise on performance of remote sensing and deep learning-based water body segmentation models," *Cybern. Syst.*, vol. 53, no. 6, pp. 581–606, Aug. 2022.

[35] X. Ma, X. Zhang, M.-O. Pun, and M. Liu, "A multilevel multimodal fusion transformer for remote sensing semantic segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5403215.

[36] X. Li, F. Xu, X. Yong, D. Chen, R. Xia, B. Ye, H. Gao, Z. Chen, and X. Lyu, "SSCNet: A spectrum-space collaborative network for semantic segmentation of remote sensing images," *Remote Sens.*, vol. 15, no. 23, p. 5610, Dec. 2023.

[37] R. Grzeszick, A. Plinge, and G. A. Fink, "Bag-of-features methods for acoustic event detection and classification," *IEEE/ACM Trans. Audio, Speech, Languages Process.*, vol. 25, no. 6, pp. 1242–1252, Jun. 2017.

[38] S. Martin, "Sequential Bayesian inference models for multiple object classification," in *Proc. 14th Int. Conf. Inf. Fusion*, Jul. 2011, pp. 1–6.

[39] X. Bian, C. Chen, L. Tian, and Q. Du, "Fusing local and global features for high-resolution scene classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 6, pp. 2889–2901, Jun. 2017.

[40] L. Bo and C. Sminchisescu, "Efficient match kernel between sets of features for visual recognition," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 22, 2009, pp. 35–44.

[41] S. C. Wong, V. Stamatescu, A. Gatt, D. Kearney, I. Lee, and M. D. McDonnell, "Track everything: Limiting prior knowledge in online multi-object recognition," *IEEE Trans. Image Process.*, vol. 26, no. 10, pp. 4669–4683, Oct. 2017.

[42] G. Sumbul, R. G. Cinbis, and S. Aksoy, "Multisource region attention network for fine-grained object recognition in remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4929–4937, Jul. 2019.

[43] H. Zhang, S.-L. Huang, and E. E. Kuruoglu, "HGR correlation pooling fusion framework for recognition and classification in multimodal remote sensing data," *Remote Sens.*, vol. 16, no. 10, p. 1708, May 2024.

[44] L. Bo and C. Sminchisescu, "Twin Gaussian processes for structured prediction," *Int. J. Comput. Vis.*, vol. 87, nos. 1–2, pp. 28–52, Mar. 2010.

[45] A. Ahmed, A. Jalal, and K. Kim, "A novel statistical method for scene classification based on multi-object categorization and logistic regression," *Sensors*, vol. 20, no. 14, p. 3871, Jul. 2020.

[46] S. Ansith and A. A. Bini, "Land use classification of high resolution remote sensing images using an encoder based modified GAN architecture," *Displays*, vol. 74, Sep. 2022, Art. no. 102229.

[47] Y.-P. Huang, L. Sithole, and T.-T. Lee, "Structure from motion technique for scene detection using autonomous drone navigation," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 49, no. 12, pp. 2559–2570, Dec. 2019.

[48] Y. Zheng, S. Liu, H. Chen, and L. Bruzzone, "Hybrid FusionNet: A hybrid feature fusion framework for multisource high-resolution remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5401714.

[49] P. Sidike, V. Sagan, M. Qumsiyeh, M. Maimaitijiang, A. Essa, and V. Asari, "Adaptive trigonometric transformation function with image contrast and color enhancement: Application to unmanned aerial system imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 3, pp. 404–408, Mar. 2018.

[50] S. Wu, M. Li, M. Zhang, K. Xu, and J. Cao, "Single base station hybrid TOA/AOD/AOA localization algorithms with the synchronization error in dense multipath environment," *EURASIP J. Wireless Commun. Netw.*, vol. 2022, no. 1, Dec. 2022, Art. no. 4, doi: 10.1186/s13638-021-02082-3.

[51] S. Wu, S. Zhang, D. Xu, and D. Huang, "A weighting localization algorithm with LOS and one-bound NLOS identification in multipath environments," *J. Inf. Sci. Eng.*, vol. 35, no. 6, pp. 1209–1222, 2019, doi: 10.6688/JISE.201911_35(6).0003.

[52] K. Liu, G. Nie, S. Jiao, B. Gao, H. Ma, J. Fu, J. Mu, and G. Wu, "Research on fault diagnosis method of vehicle cable terminal based on time series segmentation for graph neural network model," *Measurement*, vol. 237, Sep. 2024, Art. no. 114999, doi: 10.1016/j.measurement.2024.114999.

[53] S. Agarwal, "Image processing image enhancement and edge detection techniques," Univ. Inf. Technol. Waknaghat, Himachal Pradesh, India, Rep., 2018.

[54] L. Yin, L. Wang, J. Li, S. Lu, J. Tian, Z. Yin, S. Liu, and W. Zheng, "YOLOV4_CSPBi: Enhanced land target detection model," *Land*, vol. 12, no. 9, p. 1813, Sep. 2023, doi: 10.3390/land12091813.

[55] Y. Shi, J. Xi, D. Hu, Z. Cai, and K. Xu, "RayMVSNet++: Learning ray-based 1D implicit fields for accurate multi-view stereo," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 11, pp. 13666–13682, Nov. 2023, doi: 10.1109/TPAMI.2023.3296163.

[56] J. Zhao, D. Song, B. Zhu, Z. Sun, J. Han, and Y. Sun, "A human-like trajectory planning method on a curve based on the driver preview mechanism," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 11, pp. 11682–11698, Nov. 2023, doi: 10.1109/TITS.2023.3285430.

[57] S. Hatwar and A. Wanare, "GMM based image segmentation and analysis of image restoration tecniques," *Int. J. Comput. Appl.*, vol. 109, no. 16, pp. 45–50, Jan. 2015.

[58] V. Rathinam, R. Sasireka, and K. Valarmathi, "An adaptive fuzzy C-means segmentation and deep learning model for efficient mammogram classification using VGG-net," *Biomed. Signal Process. Control*, vol. 88, Feb. 2024, Art. no. 105617.

[59] B. Zhu, Y. Sun, J. Zhao, J. Han, P. Zhang, and T. Fan, "A critical scenario search method for intelligent vehicle testing based on the social cognitive optimization algorithm," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 8, pp. 7974–7986, Aug. 2023, doi: 10.1109/TITS.2023.3268324.

[60] Z. Feng, W. Ding, J. Cao, C. Sun, X. Shen, and H. Wang, "Adaptive FCM clustering algorithm based on twin multiple population genetic evolution," in *Proc. IEEE 1st Int. Conf. Digit. Twins Parallel Intell. (DTPI)*, Aug. 2021, pp. 438–441.

[61] Y. Shao, J. C.-W. Lin, G. Srivastava, A. Jolfaei, D. Guo, and Y. Hu, "Self-attention-based conditional random fields latent variables model for sequence labeling," *Pattern Recognit. Lett.*, vol. 145, pp. 157–164, May 2021.

[62] S. Li, J. Chen, W. Peng, X. Shi, and W. Bu, "A vehicle detection method based on disparity segmentation," *Multimedia Tools Appl.*, vol. 82, no. 13, pp. 19643–19655, May 2023, doi: 10.1007/s11042-023-14360-x.

[63] Z. Li, Y. Wang, R. Zhang, F. Ding, C. Wei, and J.-G. Lu, "A LiDAR-OpenStreetMap matching method for vehicle global position initialization based on boundary directional feature extraction," *IEEE Trans. Intell. Vehicles*, early access, Apr. 24, 2024, doi: 10.1109/TIV.2024.3393229.

[64] F. An, J. Wang, and R. Liu, "Road traffic sign recognition algorithm based on cascade attention-modulation fusion mechanism," *IEEE Trans. Intell. Transp. Syst.*, early access, Aug. 14, 2024, doi: 10.1109/TITS.2024.3439699.

[65] J. C.-W. Lin, Y. Shao, J. Zhang, and U. Yun, "Enhanced sequence labeling based on latent variable conditional random fields," *Neurocomputing*, vol. 403, pp. 431–440, Aug. 2020.

[66] A. W. Setiawan, "Image segmentation metrics in skin lesion: Accuracy, sensitivity, specificity, dice coefficient, Jaccard index, and Matthews correlation coefficient," in *Proc. Int. Conf. Comput. Eng., Netw., Intell. Multimedia (CENIM)*, Nov. 2020, pp. 97–102.

[67] M. Zhou, L. Chen, X. Wei, X. Liao, Q. Mao, H. Wang, H. Pu, J. Luo, T. Xiang, and B. Fang, "Perception-oriented U-shaped transformer network for 360-degree no-reference image quality assessment," *IEEE Trans. Broadcast.*, vol. 69, no. 2, pp. 396–405, Jun. 2023, doi: 10.1109/TBC.2022.3231101.

[68] Y. Jia, W. Yu, G. Chen, and L. Zhao, "Nighttime road scene image enhancement based on cycle-consistent generative adversarial network," *Sci. Rep.*, vol. 14, no. 1, Jun. 2024, Art. no. 14375, doi: 10.1038/s41598-024-65270-3.

[69] S. He, H. Luo, W. Jiang, X. Jiang, and H. Ding, "VGSG: Vision-guided semantic-group network for text-based person search," *IEEE Trans. Image Process.*, vol. 33, pp. 163–176, 2024, doi: 10.1109/TIP.2023.3337653.

[70] T. Guo, H. Yuan, R. Hamzaoui, X. Wang, and L. Wang, "Dependence-based coarse-to-fine approach for reducing distortion accumulation in G-PCC attribute compression," *IEEE Trans. Ind. Informat.*, vol. 20, no. 9, pp. 11393–11403, Sep. 2024, doi: 10.1109/TII.2024.3403262.

[71] S. He, W. Chen, K. Wang, H. Luo, F. Wang, W. Jiang, and H. Ding, "Region generation and assessment network for occluded person re-identification," *IEEE Trans. Inf. Forensics Security*, vol. 19, pp. 120–132, 2024, doi: 10.1109/TIFS.2023.3318956.

[72] Y. Liu and Y. Zhao, "A blockchain-enabled framework for vehicular data sensing: Enhancing information freshness," *IEEE Trans. Veh. Technol.*, early access, Jun. 21, 2024, doi: 10.1109/TVT.2024.3417689.

[73] J. Wu, Y. Wang, and C. Yin, "Curvilinear multilane merging and platooning with bounded control in curved road coordinates," *IEEE Trans. Veh. Technol.*, vol. 71, no. 2, pp. 1237–1252, Feb. 2022, doi: 10.1109/TVT.2021.3131751.

[74] Z. Zhou, Y. Wang, R. Liu, C. Wei, H. Du, and C. Yin, "Short-term lateral behavior reasoning for target vehicles considering driver preview characteristic," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 8, pp. 11801–11810, Aug. 2022, doi: 10.1109/TITS.2021.3107310.

[75] W. Guo, G. Xu, B. Liu, and Y. Wang, "Hyperspectral image classification using CNN-enhanced multi-level Haar wavelet features fusion network," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[76] S. Yu, D. Guan, Z. Gu, J. Guo, Z. Liu, and Y. Liu, "Radar target complex high-resolution range profile modulation by external time coding metasurface," *IEEE Trans. Microw. Theory Techn.*, early access, Apr. 16, 2024, doi: 10.1109/TMTT.2024.3385421.

[77] Z. Xiao, H. Fang, H. Jiang, J. Bai, V. Havyarimana, H. Chen, and L. Jiao, "Understanding private car aggregation effect via spatio-temporal analysis of trajectory data," *IEEE Trans. Cybern.*, vol. 53, no. 4, pp. 2346–2357, Apr. 2023, doi: 10.1109/TCYB.2021.3117705.

[78] Y. Ding, W. Zhang, X. Zhou, Q. Liao, Q. Luo, and L. M. Ni, "FraudTrip: Taxi fraudulent trip detection from corresponding trajectories," *IEEE Internet Things J.*, vol. 8, no. 16, pp. 12505–12517, Aug. 2021, doi: 10.1109/JIOT.2020.3019398.

[79] D. Cheng, L. Chen, C. Lv, L. Guo, and Q. Kou, "Light-guided and cross-fusion U-Net for anti-illumination image super-resolution," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 12, pp. 8436–8449, Dec. 2022, doi: 10.1109/TCSVT.2022.3194169.

[80] J. Gong, Y. Liu, T. Li, H. Chai, X. Wang, J. Feng, C. Deng, D. Jin, and Y. Li, "Empowering spatial knowledge graph for mobile traffic prediction," presented at the SIGSPATIAL, New York, NY, USA, Nov. 2023, doi: 10.1145/3589132.3625569.

[81] M. Al Tobi, G. Bevan, P. Wallace, D. Harrison, and K. E. Okedu, "Using MLP-GABP and SVM with wavelet packet transform-based feature extraction for fault diagnosis of a centrifugal pump," *Energy Sci. Eng.*, vol. 10, no. 6, pp. 1826–1839, 2022.

[82] Y. Xi, T. Li, H. Wang, Y. Li, S. Tarkoma, and P. Hui, "Beyond the first law of geography: Learning representations of satellite imagery by leveraging point-of-interests," presented at the WWW, New York, NY, USA, Apr. 2022, doi: 10.1145/3485447.3512149.

[83] J. Xing, H. Yuan, R. Hamzaoui, H. Liu, and J. Hou, "GQE-Net: A graph-based quality enhancement network for point cloud color attribute," *IEEE Trans. Image Process.*, vol. 32, pp. 6303–6317, 2023, doi: 10.1109/TIP.2023.3330086.

[84] Y. Sun, Z. Peng, J. Hu, and B. K. Ghosh, "Event-triggered critic learning impedance control of lower limb exoskeleton robots in interactive environments," *Neurocomputing*, vol. 564, Jan. 2024, Art. no. 126963, doi: 10.1016/j.neucom.2023.126963.

[85] Z. Ahmad, A. Rai, A. S. Maliuk, and J.-M. Kim, "Discriminant feature extraction for centrifugal pump fault diagnosis," *IEEE Access*, vol. 8, pp. 165512–165528, 2020.

[86] J. Chen, Q. Wang, W. Peng, H. Xu, X. Li, and W. Xu, "Disparity-based multiscale fusion network for transportation detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 10, pp. 18855–18863, Oct. 2022, doi: 10.1109/TITS.2022.3161977.

[87] J. Yang, C. Wu, B. Du, and L. Zhang, "Enhanced multiscale feature fusion network for HSI classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10328–10347, Dec. 2021.

[88] M. Hanzla, M. O. Yusuf, N. Al Mudawi, T. Sadiq, N. A. Almujally, H. Rahman, A. Alazeb, and A. Algarni, "Vehicle recognition pipeline via DeepSort on aerial image datasets," *Frontiers Neurorobot.*, vol. 18, Aug. 2024, Art. no. 1430155, doi: 10.3389/fnbot.2024.1430155.

[89] A. Nadeem, A. Jalal, and K. Kim, "Accurate physical activity recognition using multidimensional features and Markov model for smart health fitness," *Symmetry*, vol. 12, no. 11, p. 1766, Oct. 2020, doi: 10.3390/sym12111766.

[90] M. S. Alshehri, M. O. Yusuf, and M. Hanzla, "Unmanned aerial vehicle detection and tracking using image segmentation and Bayesian filtering," in *Proc. 4th Interdiscipl. Conf. Electrics Comput. (INTCEC)*, vol. 13, Chicago, IL, USA, Jun. 2024, pp. 1–6, doi: 10.1109/intcec61833.2024.10602916.

[91] J. Chen, M. Xu, W. Xu, D. Li, W. Peng, and H. Xu, "A flow feedback traffic prediction based on visual quantified features," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 9, pp. 10067–10075, Sep. 2023, doi: 10.1109/TITS.2023.3269794.

[92] M. Hanzla, S. Ali, and A. Jalal, "Smart traffic monitoring through drone images via YOLOv5 and Kalman filter," in *Proc. 5th Int. Conf. Adv. Comput. Sci. (ICACS)*, Lahore, Pakistan, Feb. 2024, pp. 1–8, doi: 10.1109/icacs60934.2024.10473259.

[93] M. O. Yusuf, M. Hanzla, H. Rahman, T. Sadiq, N. A. Mudawi, N. A. Almujally, and A. Algarni, "Enhancing vehicle detection and tracking in UAV imagery: A pixel labeling and particle filter approach," *IEEE Access*, vol. 12, pp. 72896–72911, 2024, doi: 10.1109/ACCESS.2024.3401253.

[94] A. Jalal, A. Ahmed, A. A. Rafique, and K. Kim, "Scene semantic recognition based on modified fuzzy C-mean and maximum entropy using object-to-object relations," *IEEE Access*, vol. 9, pp. 27758–27772, 2021, doi: 10.1109/ACCESS.2021.3058986.

[95] Q. Chen and G. Pan, "A structure-self-organizing DBN for image recognition," *Neural Comput. Appl.*, vol. 33, no. 3, pp. 877–886, Feb. 2021.

[96] T. Yang and D. L. Silver, "The disadvantage of CNN versus DBN image classification under adversarial conditions," in *Proc. Can. Conf. AI*, 2021, pp. 109–115.

[97] P. Dou, H. Shen, Z. Li, X. Guan, and W. Huang, "Remote sensing image classification using deep–shallow learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 3070–3083, 2021.

[98] O. Almomani, "A feature selection model for network intrusion detection system based on PSO, GWO, FFA and GA algorithms," *Symmetry*, vol. 12, no. 6, p. 1046, Jun. 2020.

[99] M. O. Yusuf, M. Hanzla, N. Al Mudawi, T. Sadiq, B. Alabdullah, H. Rahman, and A. Algarni, "Target detection and classification via EfficientDet and CNN over unmanned aerial vehicles," *Frontiers Neurorobot.*, vol. 18, Aug. 2024, Art. no. 1448538, doi: 10.3389/fnbot.2024.1448538.

[100] J. Chen, Q. Wang, H. H. Cheng, W. Peng, and W. Xu, "A review of vision-based traffic semantic understanding in ITSs," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 11, pp. 19954–19979, Nov. 2022, doi: 10.1109/TITS.2022.3182410.

[101] J. S. Shukla, K. Rastogi, H. Patel, G. Jain, and S. Sharma, "Bag of visual words methodology in remote sensing—A review," in *Proc. Int. e-Conf. Intell. Syst. Signal Process.*, 2020, pp. 475–486.

[102] G. Cheng, X. Xie, J. Han, L. Guo, and G.-S. Xia, "Remote sensing image scene classification meets deep learning: Challenges, methods, benchmarks, and opportunities," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 3735–3756, 2020.

[103] H. Xie, Y. Chen, and P. Ghamisi, "Remote sensing image scene classification via label augmentation and intra-class constraint," *Remote Sens.*, vol. 13, no. 13, p. 2566, Jun. 2021.

[104] J. Kim and M. Chi, "SAFFNet: Self-attention-based feature fusion network for remote sensing few-shot scene classification," *Remote Sens.*, vol. 13, no. 13, p. 2532, Jun. 2021.

[105] Y. Yu and F. Liu, "A two-stream deep fusion framework for high-resolution aerial scene classification," *Comput. Intell. Neurosci.*, vol. 2018, no. 1, pp. 1–13, 2018.

[106] W. Wang, Y. Sun, J. Li, and X. Wang, "Frequency and spatial based multi-layer context network (FSCNet) for remote sensing scene classification," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 128, Apr. 2024, Art. no. 103781.

[107] W. Han, R. Feng, L. Wang, and Y. Cheng, "A semi-supervised generative framework with deep learning features for high-resolution remote sensing image scene classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 145, pp. 23–43, Nov. 2018.

[108] G. Cheng, C. Yang, X. Yao, L. Guo, and J. Han, "When deep learning meets metric learning: Remote sensing image scene classification via learning discriminative CNNs," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2811–2821, May 2018.
[109] X. Li, C. Wen, Y. Hu, and N. Zhou, "RS-CLIP: Zero shot remote sensing scene classification via contrastive vision-language supervision," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 124, Nov. 2023, Art. no. 103497.
[110] C. Sitaula, S. Kc, and J. Aryal, "Enhanced multi-level features for very high resolution remote sensing scene classification," *Neural Comput. Appl.*, vol. 36, no. 13, pp. 7071–7083, May 2024.
[111] X. Liu, Y. Zhou, J. Zhao, R. Yao, B. Liu, and Y. Zheng, "Siamese convolutional neural networks for remote sensing scene classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 8, pp. 1200–1204, Aug. 2019.

MUHAMMAD WAQAS AHMED received the M.S. degree in computer sciences from COMSATS. He is currently pursuing the Ph.D. degree in computer science with Air University, Islamabad, Pakistan. His research interests include artificial intelligence, computer vision, machine learning algorithms, deep learning, image and video processing, and intelligent systems.

ABDULLAH ALSHAHRANI received the Graduate degree in computer science from King Khalid University, in 2008, the M.Sc. degree in computer science from La Trobe University, Melbourne, Australia, in 2010, and the Ph.D. degree from The Catholic University of America, USA, in 2018. He is currently an Assistant Professor with the Department of Computer Science and Artificial Intelligence, College of Computer Science and Engineering, University of Jeddah, Saudi Arabia. His research interests include wireless sensor networks, network security, parallel computing, smart homes systems, the IoT, and data science.

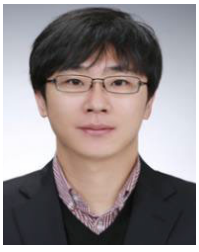
ABRAR ALMJALLY received the bachelor's degree from King Saud University, in 2007, the master's degree in information management security from Syracuse University, USA, in 2014, and the Ph.D. degree in informatics from the University of Sussex, Brighton, U.K. She is currently a Faculty Member with the Information Technology Department, Al Imam Mohammad Ibn Saud Islamic University (IMSIU). Her research interests include computing education, the Internet of Things, education technology, machine learning, and artificial intelligence.

NAIF AL MUDAWI received the master's degree in computer science from Australian La Trobe University, in 2011, and the Ph.D. degree from the College of Engineering and Informatics, University of Sussex, Brighton, U.K., in 2018. He is currently an Assistant Professor with the Department of Computer Science and Information System, Najran University. During his academic journey to obtain the master's degree, he was a member of Australian Computer Science Committee. He has many published research and scientific articles in many prestigious journals in various disciplines of computer science.

ASAAD ALGARNI received the Ph.D. degree in software engineering from North Dakota State University, USA. He is currently an Assistant Professor with the Department of Computer Sciences, College of Computing and Information Technology, Northern Borders University, Saudi Arabia. His research interests include software engineering, computer vision applications, and machine learning.

KHALED AL NOWAISER received the Ph.D. degree in computer science from Glasgow University, Scotland. He is currently an Assistant Professor with the Computer Engineering Department, Prince Sattam Bin Abdulaziz University, Saudi Arabia. His research interests include computer vision, optimization techniques, and performance enhancement.

AHMAD JALAL received the Ph.D. degree from the Department of Biomedical Engineering, Kyung Hee University, Republic of Korea. He is currently an Associate Professor with the Department of Computer Science and Engineering, Air University, Pakistan. He was a Postdoctoral Research Fellowship with POSTECH. His research interests include multimedia contents and artificial intelligence.



JEONGMIN PARK received the Ph.D. degree from the College of Information and Communication Engineering, Sungkyunkwan University, in 2009. He is currently an Associate Professor with the Department of Computer Engineering, Tech University of Korea, South Korea. Before joining the Tech University of Korea, in 2014, he was a Senior Researcher with the Electronics and Telecommunications Research Institute (ETRI) and a Research Professor with Sungkyunkwan University, South Korea. His research interests include high-reliable autonomic computing mechanism and human-oriented interaction systems.