IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 10 June 2024, accepted 28 June 2024, date of publication 3 July 2024, date of current version 12 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3422368

RESEARCH ARTICLE

A Novel Deep Learning Framework With Meta-Heuristic Feature Selection for Enhanced Remote Sensing Image Classification

BILAL AHMED[1], TALLHA AKRAM[1], SYED RAMEEZ NAQVI[1], ANAS ALSUHAIBANI[2], YOUSSEF N. ALTHERWY[2], AND USMAN MASUD[3]

[1]Department of Electrical and Computer Engineering, COMSATS University Islamabad, Wah Campus, Rawalpindi 47040, Pakistan
[2]Department of Information Systems, College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia
[3]Department of Electrical Communication Engineering, University of Kassel, 34127 Kassel, Germany

Corresponding author: Youssef N. Altherwy (y.altherwy@psau.edu.sa)

This work was supported by Prince Sattam Bin Abdulaziz University under Project PSAU/2024/R/1445.

ABSTRACT We propose a novel deep learning architecture, called XcelNet17, for image classification in remote sensing. Comprising fourteen convolutional and three fully connected layers, XcelNet17 outperforms several benchmark architectures available in the literature in terms of classification accuracy. Additionally, we present BA-ABC, a new hybrid feature selection algorithm that inherits the strengths of the Bat Algorithm (BA) and the Artificial Bee Colony (ABC) algorithm. Together these contributions significantly enhance the performance and accuracy of remote sensing image classification tasks. The proposed framework is thoroughly trained and verified using five benchmark datasets typically used for remote sensing image classification, namely AID, RSSCN7, SIRI-WHU, UC Merced, and WHU RS-19. Our simulation results suggest that in terms of classification accuracy, XcelNet17 outperforms most of the well established networks including AlexNet, VGG16, VGG19, ResNet50, and DarkNet19 by obtaining accuracy values in the range of 94.6% and 99.9%. Furthermore, the proposed features selection method, when integrated with XcelNet17, yields much improved classification accuracy in comparison to various benchmarks including WOA, GWO, BA, ABC, and ACO algorithms. For example, an 8% superior performance on WHU-RS19 dataset has been observed. The attained results are further validated by an in-depth statistical analysis.

INDEX TERMS Artificial bee colony, bat algorithm, bio-inspired feature selection, CNN architecture, image classification, remote sensing.

I. INTRODUCTION

Remote Sensing (RS) is the acquisition of data about a phenomenon or object without direct physical contact. It relies on sensors to detect and capture electromagnetic radiation emitted or reflected by the target. RS is mainly classified into two categories: active and passive. While the former requires the source, such as a radar or a lidar, to emit its own energy to illuminate the target, the passive relies on sunlight to acquire the target's characteristics, such as color and temperature. Remote sensing finds extensive applications across a diverse array of disciplines such as agriculture, forestry, meteorology, geology, and environmental monitoring, serving as a pivotal tool in the acquisition and analysis of data pertinent to these fields [1].

Image classification, which is a common task in computer vision for assigning a label or class to an image based on its visual content, becomes non-trivial in RS. One of the many reasons behind this predicament is the complexity and diversity of aerial/satellite images, which are affected

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

by parameters like resolution, illumination, noise, occlusion, and seasonal changes [2]. The traditional approaches for RS image classification rely either on handcrafted features, such as color, texture, shape, and edge, or on machine learning algorithms, like decision trees, neural networks, and support vector machines. The latter have shown good performance on small-scale and low-resolution datasets, but they have limitations in capturing high-level semantic information and dealing with high-resolution and large-scale datasets that have become more available in recent years [3]. Feature selection (FS) methods are a critical step in image classification and have shown effectiveness in various scenarios. This is particularly true for RS images, which often contain a mix of spatial, spectral, tactile, and contextual information; these features are not always relevant, and can be redundant and noisy. Therefore, FS methods usually aim to reduce feature dimensionality, boost classification efficacy, and enhance interpretability of the results [4]. There are different types of FS methods such as wrapper, filter, embedded, and hybrid [5], which use different criteria and strategies to evaluate and select the optimal subset of features. While, some of the common criteria are correlation, mutual information, entropy, Fisher score, and sparsity; the more recent strategies employed in FS are genetic algorithm [6], backward elimination, and forward selection [7]. FS techniques can also be classified into supervised, unsupervised, and semi-supervised methods [8], depending on whether they use the class labels or not.

The nature-inspired FS methods have recently received a lot of attention and praise in this context. Their inherent ability to mimic the behavior of various organisms to search for optimal solutions, has made them the de-facto choice for FS in RS image classification in recent years. Some common nature-inspired algorithms, such as ant colony optimization (ACO) [9], artificial bee colony (ABC) [10], bat algorithm (BA) [11], firefly algorithm (FA) [12], grey wolf optimization (GWO) [13], particle swarm optimization (PSO) [14], and whale optimization algorithm (WOA) [15], simulate the behavior of ants, bees, bats, fireflies, grey wolves, birds and whales respectively. The most recent of the lot to gain fame in this context are the BA and ABC algorithms. The former aims to find the loudest solution (feature subset) by adjusting the frequency and velocity of each bat (individual) according to the current best solution and a randomization factor. ABC, on the other hand, aims to find the optimal solution by applying three processes: employee bees, onlookers, and scouts. The employee bees exploit the current solutions (feature subsets) by generating new ones from which the onlookers select the best solutions based on a probabilistic rule and generate new ones (feature subsets) in their neighborhood. The scouts abandon the worst solutions and randomly generate new ones in the search space. Certain distinct capabilities of each of these two algorithms give them an edge over one another, and it rather becomes needless to advocate for one. For example, the BAT algorithm, due to its echolocation-inspired mechanism, is renowned for its effective exploration of the search space, the ABC algorithm, on the other hand, enjoys better exploitation of promising regions in the search space. We believe that hybridizing the two algorithms will not only promise us improved exploration but will allow us to strike a balance between exploration and exploitation - resulting in a more robust FS method, which is the primary contribution of this work.

In recent years, significant advancements have been made in RS image classification through the utilization of Deep Convolutional Neural Network (DCNN) models, facilitated by advancements in deep learning. Convolution operations are adept at capturing local image information effectively. Studies such as [16] have demonstrated that convolutional layers of varying depths can extract distinct features. To incorporate global features, neural networks leveraging convolution operations require the stacking of multiple layers. AlexNet [17], for instance, is the first deep CNN model applied for image classification and recognition. Similarly, ResNet [18] aims to deepen CNNs for enhanced training ease. Motivated by the same, in this work we propose a novel deep learning architecture called XcelNet17, which comprises fourteen convolutional and three fully connected layers. When integrated with the proposed BA-ABC algorithm for FS, and thoroughly trained using five benchmark datasets including AID [19], RSSCN7 [20], SIRI-WHU Dataset [21], UC Merced [22], and WHU-RS19 datasets [23], [24], XcelNet17 outperforms most of the existing models such as AlexNet [17], VGG16 [25], VGG19 [25], ResNet50 [18], and DarkNet19 [26] by obtaining accuracy values in the range of 94.6% and 99.9%. On the other hand, the BA-ABC algorithm, when integrated with XcelNet17, yields much improved classification accuracy in comparison to various benchmarks including WOA, GWO, BA, ABC, and ACO algorithms.

The rest of the manuscript is organized as follows: Sect. II and Sect. III provide an overview of related work and enumerate major contributions respectively. Sect. IV provides a brief overview of the datasets used, followed by the proposed architecture XcelNet17. It also presents details of the proposed BA-ABC hybrid FS algorithm. Sect. V presents the simulation results and statistical and convergence curve analyses. Finally, Sect. VI concludes the manuscript.

II. RELATED WORK

In the past decade, significant efforts have been devoted to developing numerous methods for categorizing scenes using satellite or aerial images. The process of image classification typically involves two stages: (i) feature extraction and (ii) classification. Various conventional as well as deep learning approaches have been developed. A new architecture, Hydra is introduced in [27], which creates an initial coarsely optimized CNN while fine-tuning of obtained weights is performed several times with various augmentation techniques, class weights, and crop styles. An ensemble of CNNs is created using ResNet & DenseNet and reduction in training

time is the main advantage with state-of-the-art performance. The efficacy of deep learning based discriminative CNNs (ADSSM, CNNs-WD and Hydra) is reviewed in [28]. More than two dozen algorithms are evaluated on three benchmark data sets.

A novel recurrent attention structure, inspired by the human visual system is proposed in [29]. The scene is classified using the attention recurrent convolutional network (ARCNet). This method empowers classifiers to focus on key areas, accelerates convergence rate, and improves accuracy. Experiments depict that ARCNet outperforms most of the methods available in the literature, and this research portrays the effectiveness of attention mechanism for RS scene classification. A Hierarchical Wasserstein Distance (HWD)-based CNN is introduced in [30] which models interclass relationship as an arrangement of categories. Information from three CNNs is merged to form category trees and distance among two distributions is measured in data space, which is hierarchically organized via HWD. A multiscale CNN (MCNN) is developed for object scale variation problem in [31]. This network trains CNN using multiple scales (F-net and V-net). For the training stage, a fixed scale is used for the F-net, while a random scale is used for the V-net, which is altered in every n^{th} iteration. For performance enhancement, MCNN is added on with a similarity measure layer. Fang et al. [32] have used space–frequency joint representation for RS scene classification problems. In space-domain, the relation between various local areas and the interactions between various global features are encoded. Band-pass filter network is used to extract statistical and stable frequency-domain features. Circular convolutional module (CCM) fuses the attained space and frequency-domain features. In fact, CCM represents relation of the features of diverse domains and obtains a discriminative and robust representation.

Wang et al. [33] have proposed a Representation-enhanced Status Replay Network (RSRNet) to address the issues of saturated network performance and the limited effective use of complementary data. These are caused by representation and classifier biases accumulation and the imbalance of fusion information among multisource images. To address these issues, First, dual augmentation, including modal and semantic augmentation, enhances the transferability and discreteness of feature representation, reducing representation bias. Then, a status replay strategy (SRS) is introduced to mitigate classifier bias and maintain decision boundary stability. Finally, a novel cross-modal interactive fusion (CMIF) method is employed to enhance the interactivity of modal fusion, optimizing the parameters of different branches by integrating multisource information.

FS has been a vibrant and productive domain within the research communities of machine learning, data mining, pattern recognition, and statistics [34], [35]. The fundamental goal of FS is to streamline data by retaining only those features that contribute positively to the accuracy of classification. This approach has demonstrated effectiveness both in theoretical frameworks and practical applications [36].

Researchers employ various FS methods; some of those favor conventional approaches such as chi-square and information gain [46], [47], [48]. On the other hand, heuristic methods like genetic algorithms (GA) [49], ACO [37], and memetic FS [39] are adopted by others to address specific challenges like noisy data, spam emails, and binary variables. Several FS techniques can also be employed to enhance the typical performance of algorithms, as described in [50]. Various enhanced techniques have been proposed to achieve better results compared to the original ones. For instance, the authors in [38] and [51] have introduced an improved version of GWO called Random-GWO. Similarly, [52] and [53] are improved versions of the original Salp Swarm and Chaotic Dragonfly algorithms, respectively. Furthermore, the literature presents enhanced methodologies like the improved GA [54] and the improved PSO, where the authors have exploited parallelism to ensure efficient computing.

Rodrigues et al. [55] have introduced a wrapper FS method that combines BA and optimum path forest (OPF) algorithms. This approach conceptualizes FS as a binary optimization method. The authors have used six datasets for experimentation, revealing that the suggested technique yields more compact and statistically significant sets with improvement in classification accuracy in some instances. The binary version of BA (BBA) has been designed to identify the most relevant features within a search space [40]. BBA assigns a set of binary coordinates to each bat, indicating whether a feature is included in the final set or otherwise. By combining the capabilities of BA and OPF, it seeks to identify a feature set that maximizes accuracy in validating sets. The proposed technique has demonstrated superiority over various established methods such as PSO, GSA, and FFA. Following in the same line, Manchala et al. introduce a binary version of the ABC algorithm, and apply it to the problem of intrusion detection [56]. They exploit a hybrid fitness function that combines information gain and correlation-based FS to measure relevance and redundancy of the features. For the same application of intrusion detection, the ABC algorithm has been adopted once again in [42], and the results are compared with other methods including genetic algorithm, PSO, and ACO. The paper uses a wrapper strategy and support vector machine (SVM) classifier to evaluate the features.

In another study [41], the authors have presented a bio-inspired approach called BANB, which is a hybrid of the BA algorithm and Naïve Bayes Classifier. Experimentation with twelve benchmark datasets from various domains reveal BANB's performance, in terms of the count of the selected features, against three well-established FS techniques (PSO, GA, and GPSO). The results indicate BANB's superiority over the rest, further leading to improved classification accuracy. Similarly, Modified-BA (MBA) algorithm incorporates

TABLE 1. Overview of existing techniques.

Method	Dataset	Classes/S	Perf (%)	Ref
Hydra (ensemble of ResNet & DenseNet)	functional map of world (FMOW) NWPU-RESISC45	62 45	78.73 94.51	[27]
Discriminative CNNs	UC Merced AID NWPU-RESISC45	21 30 45	98.93 96.19 91.89	[28]
ARCNet-VGGNet16	UC Merced AID RS19 OPTIMAL-31	21 30 19 31	99.12 93.10 99.50 92.70	[29]
Hierarchical Wasserstein (HW) CNN	AID NWPU-RESISC45	30 45	93.27 94.38	[30]
Multiscale CNN (MCNN)	UC Merced AID SIRI- WHU	21 30 12	74.60 73.10 72.60	[31]
Space–Frequency joint representation SVM	Sydney UC-Merced WHU-RS19 AID	7 21 19 30	97.36 97.21 97.48 96.18	[32]
Representation-enhanced Status Replay Network (RSRNet)	Three Datasets	-	-	[33]
Ant Colony Optimization (ACO) for Feature Selection	UCI Datasets	2-26	94.80	[37]
Improved Grey Wolf Optimizer (IGWO)	CEC2018 benchmark suite	50	94.20	[38]
Bare-bones Particle Swarm Optimization for Feature Selection	Ionosphere Sonar	27 24	96.21 96.08	[39]
Binary Bat Algorithm (BBA) for Feature Selection	Breast Cancer Australian German Numer DNA Mushrooms1	2 2 3 2 2	77.25 96.31 83.02 70.24 99.95	[40]
BA with a Naive Bayes classifier (BANB)	M-of-N Exactly Exactly2 UCI datasets	1000 1000 1000 294-8124	98.90 68.80 75.80 87.90	[41]
Modified Bat Algorithm (MBA)	Wisconsin Diagnosis Breast Cancer (WDBC)	2	96.00	[42]
Nature-inspired Algos (GA,BFOA,GBF,KHA,WCA,SOSA) in scheduling tasks - A Review	-	-	-	[43]
Classical methods and Nature-Inspired Algos (NIAs) for Optimization - A Review	-	-	-	[44]
Hybrid Artificial Bee Colony (ABC) and Genetic Algorithm (GA)	UCI Datasets	2-7	86.30	[45]

a simple random sampling technique for selecting random instances from the dataset [57]. The method prioritizes the ranking of features based on their global significance to identify the most prominent features within the dataset. The chosen features are then utilized in training of the random forest classifier. In another attempt, the authors [58] have come up with a FS method for analyzing data based on ABC which can be applied to various domains. The technique utilizes a wrapper strategy and a forward selection technique in evaluating the selected features quality.

Nayak [43] have examined the incorporation of nature-inspired algorithms in scheduling tasks. This paper introduces task scheduling utilizing various types of nature-inspired algorithms such as genetic algorithms, water cycle algorithms, etc. Mandal [44] provide an overview of the latest application of nature-inspired algorithms and traditional techniques for addressing single and multi-objective optimization issues. This paper categorizes nature-inspired algorithms into three groups: swarm intelligence-based, evolutionary algorithm, and other metaheuristics. Evolutionary algorithms and swarm intelligence-based nature-inspired algorithms can be categorized into five subcategories. This study offers a thorough view for choosing the most suitable approach for addressing various optimization problems based on their level of complexity. Much closer to our proposed work is a hybrid method for FS that combines ABC and GA to exploit the advantages of both works [45]. The method uses a filter strategy and a mutual information criterion to evaluate the features. A summary of the most relevant works is presented in Table 1.

III. PROBLEM STATEMENT AND CONTRIBUTIONS

The studies summarized in section II highlight the effectiveness of DL architectures and FS techniques in improving the performance of learning algorithms. Specifically, they demonstrate how classifier accuracy can be enhanced through the elimination of insignificant features. However, it is a well-known fact that simple yet effective architectures and high-quality features contribute to accurate, comprehensible classification processes, yielding improved results. Consequently, this research aims to experimentally analyze and identify superior architectures and algorithmic techniques among both conventional and heuristic approaches.

In this research, we introduce a new architecture for image classification, XcelNet17. This architecture, despite its simplicity, exhibits outstanding performance. Exhaustive simulations are carried out for performance evaluation of XcelNet17 with state-of-the-art models like AlexNet, ResNet, VGG, and DarkNet on five popular RS datasets, which confirm its superior efficacy. In addition, an optimization algorithm called BA-ABC is proposed, which aims for optimal FS in RS image classification. Through simulation cases, we compare BA-ABC with other established methods like WOA, GWO, ACO, BA, and ABC, confirming its ability to mitigate premature convergence and improve classification accuracy significantly. The main contributions of this research work are enumerated as follows:

1) For RS image classification, a novel and cutting-edge deep-learning architecture *XcelNet17* is proposed, which addresses the problems of over-fitting, high computing cost, and the curse of dimensionality.
2) We hybridize the BA algorithm with a leadership hierarchy approach, and incorporate the upgrading strategy of an employed bee from the ABC algorithm. This integration facilitates the sharing of information among group members, enhancing exploitation while preserving the essential features of the conventional BA algorithm.
3) A chaotic mapping population initialization method is utilized in the BA-ABC algorithm, which ensures a diverse set of initial solutions.
4) We benchmark XcelNet17 and the BA-ABC algorithm on five RS yardstick datasets with diverse characteristics, and compare them against five established CNN models, and algorithms, respectively. The effectiveness is validated via comprehensive simulations using the Monte-Carlo approach. The evaluation includes ranking on accuracy, F1 score, sensitivity, precision, and specificity.

IV. MATERIALS AND METHOD

This section presents the datasets used in this research work followed by details of the proposed architecture and the FS algorithm.

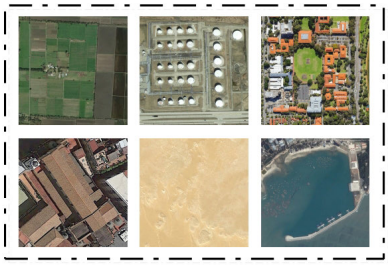


FIGURE 1. Sample images from AID dataset.

FIGURE 2. Sample images from RSSCN7 dataset.

A. DATASETS

In this study we have utilized the following datasets: AID [19], RSSCN7 [20], SIRI-WHU Dataset [21], UC Merced [22], and WHU-RS19 datasets [23], [24]. The AID dataset [19] consists of a wide range of aerial images obtained by extracting images from Google Earth. The images are then processed using RGB rendering based on actual aerial optical images. This dataset has aerial scenes of 30 types: storage tanks, viaduct, stadium, sparse residential, school, square, river, railway station, resort, pond, port, parking, playground, mountain, park, medium residential, industrial, meadow, forest, desert, farmland, dense residential, commercial, center, church, beach, bridge, baseball field, bare land, and airport. The total number of images in this dataset is 10,000 arranged in 30 different classes. The spatial resolution varies from $\approx$ 8 meters to $\approx$ 0.5 meters. The RS image interpretation experts have labeled all the images. Figure 1 shows examples of some of the classes.

The RSSCN7 dataset, as described in [20], consists of a diverse array of scene images captured in various seasons and under different weather conditions. The dataset includes seven distinct image classes with a total of 400 images in each class. The image size is 400 $\times$ 400 pixels and are distributed across four different scales. This results in 100 scenes per scale. Figure 2 illustrates examples from the dataset.

The third dataset is the SIRI-WHU [21] dataset, which consists of 2400 images. The dataset is categorised into 12 distinct scene classes with each class containing 200 images. The image resolution and size are 2m and 200 *times* 200 pixel respectively. Figure 3 shows examples from the dataset.

The UC Merced dataset [22] is the fourth dataset used in this work. It contains 2100 images distributed equally across

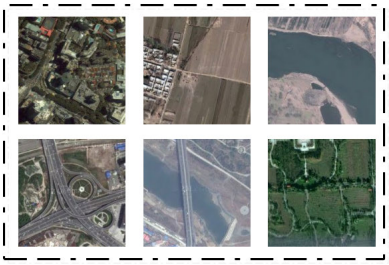


FIGURE 3. Sample images from SIRI-WHU dataset.

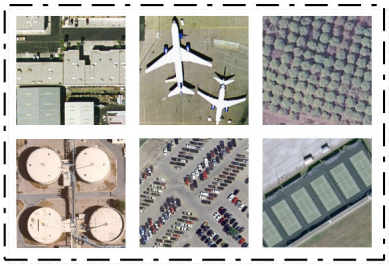


FIGURE 4. Sample images from UC Merced dataset.

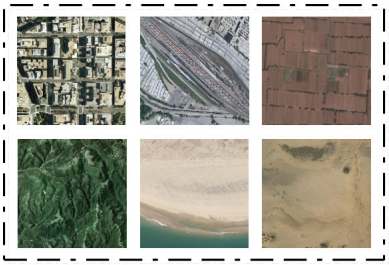


FIGURE 5. Sample images from WHU-RS19 dataset.

21 classes. These classes encompass urban, agricultural, and natural landscapes. The images in the dataset are extracted from the National Map Urban Area Imagery collection of the USGS, showcasing a variety of urban locales across the USA. The images have a pixel resolution of 0.3 meters and a size of 256×256 pixels, except for 44 images that differ in shape and dimensions from the mentioned metrics. Figure 4 provides examples from this dataset.

The final dataset is the WHU-RS19 [23], [24]. The dataset is a compilation of satellite images sourced from Google Earth, providing high-resolution visuals with a granularity of up to 0.5 meters. The dataset encompasses 19 distinct classes representing scenes captured using high-resolution satellite imagery.

These classes include airport, bridge, beach, desert, commercial area, farmland, forest, football field, industrial zone, meadow, mountainous terrain, parking area, park, port, pond, railway station, residential zone, viaduct, and river. Approximately 50 samples are available for each class. Images from the same class are sourced from diverse regions within satellite images of varying resolutions, resulting in potential differences in scales, orientations, and illuminations. Figure 5 showcases examples from the dataset.

TABLE 2. Datasets description.

Dataset	Images	Classes	Images/ class	Size	Resol
AID	10000	30	334	600 x 600	8 - 0.5
RSSCN7	2800	7	400	400 x 400	-
SIRI-WHU	2400	12	200	200 x 200	2
UC Merced	2100	21	100	256 x 256	0.3
WHU-RS19	1005	19	50	600 x 600	0.5

Pre-processing imbalanced remote sensing datasets is essential to ensure that the subsequent machine learning models perform well. In this research, we have utilized random data splitting in the initial phase, normalization, and feature selection for this purpose. Random data split ensures that no specific weightage is given to either the majority or minority classes, thus increasing the probability of a balanced sample set. We have used normalization to scale the feature values to a standard range or distribution to ensure that all features contribute equally to the model's performance. Furthermore, a novel nature-inspired feature selection algorithm, BA-ABC is used to select the most relevant features for the classification task. These steps are designed to mitigate the challenges posed by imbalanced datasets, such as the model's bias towards the majority class, and to enhance the predictive accuracy of the machine learning algorithms used for remote sensing data analysis. Table 2 presents an overview of the five datasets. In the next section, we discuss the proposed architecture - XcelNet17.

B. PROPOSED ARCHITECTURE - XCELNET17

In this study, we have developed XcelNet17, a new architecture for RS image classification. It addresses the challenge of a multi-class classification problem involving diverse categories. XcelNet17 is intended to achieve high-precision RS image classification based on extracted features and to evaluate the influence of Deep Learning models in the field of image classification. It has a total of 48 layers, with 14 layers of convolution, 3 fully connected, 3 max-pooling layers, as well as relu and cross-normalization layers. A combination of simple and residual blocks is implemented, where the residual block comprises five residual stacks. The network is structured to benefit from the strengths of both simple as well as residual themes, yielding excellent performance even with less computational complexity.

The convolution layers employ filter sizes of 3×3, 5×5, and 11×11. Utilizing diverse filter sizes enables the network to efficiently gather features compared to using a single size, thus enhancing computational efficiency. This design promotes a balanced receptive field, facilitating the collection of elements with varying sizes and complexities. Each convolution layer is configured with 96, 256, and 384 filters, in various layers. A stride value of 1 is utilized, along with padding values of 0, 2, and 1. Additionally, ReLU activation functions are applied. For XcelNet17 initialization,

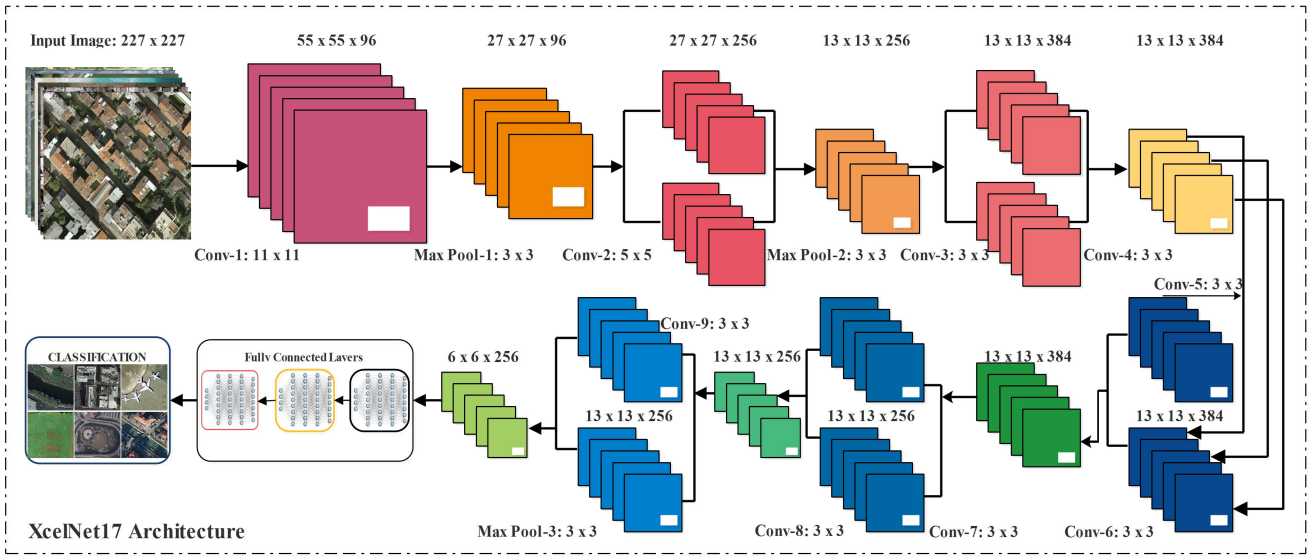


FIGURE 6. Architecture of the proposed network XcelNet17.

the He method is used as it considers the size of the previous layer in the network to maintain a variance that allows the signal to flow well in the forward and backward pass. This method scales the initial weights based on the number of input and output neurons, thus avoiding the core problems of exploding and vanishing gradients. Figure 6 depicts XcelNet17 architecture.

The XcelNet17 is trained on the five selected datasets. After network training, feature extraction is performed utilizing the trained model. The features are extracted from the fully connected layer, and are fed to the classifier. Several classifiers are applied to the extracted features and it has been observed that Cubic SVM, Medium Neural Network, and Cubic KNN yield the best classification accuracy. Utilization of multiple classifiers provides valuable insights into the effectiveness of the extracted features and the robustness of the XcelNet17 architecture. These findings underscore the adaptability and generalization capabilities of the trained model across diverse datasets and classification tasks. Furthermore, the observed high accuracy attained by the Cubic SVM, Medium Neural Network, and Cubic KNN classifiers highlight their suitability for RS image classification, affirming their effectiveness in leveraging the extracted features for accurate classification.

C. PROPOSED ALGORITHM: BA-ABC

The proposed FS algorithm, BA-ABC, enhances search efficiency by modifying solutions with inferior fitness to introduce diversity into the population. The BA-ABC algorithm is developed by combining the Bat and ABC algorithms. The former is a meta-heuristic algorithm inspired by nature. It was introduced by Yang [11] in 2010, and it draws inspiration from the bats' echolocation behavior, which they use to sense distances. During nocturnal hunting, bats emit short-duration loud sound pulses and listen for the echo rebounding from obstacles or prey. Through their specialized auditory mechanism, bats can discern the size and position of objects. The BA algorithm has found extensive use in various applications, including pattern recognition [59], and engineering optimization [60], [61].

The ABC algorithm, on the other hand, is based on swarm intelligence inspired by the honey bee's behavior [62]. The algorithm generates an initial population of N solutions (referred to as food sources), distributed randomly, with N representing the swarm size.[1] Let X_i be the i_{th} solution and n be size of the dimension. Every employee bee X_i produces a new possible solution X^* in its current neighbor as follows:

$$v_{j,k} = \phi_{j,k}(X_{j,k} - X_{j,i}) + X_{j,k} \tag{1}$$

here X_k is a random possible solution $(i \neq k)$,k depicts a randomly chosen dimension index from the set $1, 2, \ldots, n$, and ϕ has a range of $[-1, 1]$ and is randomly generated. A greedy selection method is used as the new possible solution v_i is generated. If v_i is fitness-wise superior to its parent x_i, then v_i replaces x_i; otherwise, there is no change.

Once the search process is finished for all employee bees, they do a waggle-dance to communicate food source information to the onlooker bees. The nectar information collected from all employee bees is assessed by every onlooker bee which chooses a source of food utilizing the amount of nectar it contains. This selection process follows a probabilistic approach, similar to roulette wheel selection. The process is detailed as follows:

$$p_i = \frac{fit_i}{\sum_{j=1}^{N} fit_j} \tag{2}$$

[1] Refer to Table 3 for the algorithmic parameters description.

TABLE 3. Algorithmic parameters' description.

Parameter	Description
N	Population size
it_{max}	Max iterations
i	Current iteration, in the range [1, N]
X_b	Global best position (solution)
X*	Current best position (solution)
X_{ij}	Neighbor Search Agent
X_i	Position in iteration i
v_i	Velocity in iteration i
f_i	Pulse frequency in iteration i and has range from f_{min} to f_{max}
r_i	Pulse rate of agent (bat) i
A_i	Loudness of agent (bat) i
α	Constant parameter in [0, 1] range for updating loudness A
γ	Constant parameter in [0, 1] range for updating pulse rate r
$[\delta_u, \ldots, \delta_l]$	Upper and lower bounds
$a, b, r, r_1, r_2, \beta, \phi$	Random Numbers
n	Number of dimensions, $1, \ldots, d$

where fit_i represents the fitness of the i_{th} swarm solution. As observed, the higher the quality of the solution i, the greater the probability of selecting the i_{th} food source.

A food source is abandoned if it can't be enhanced for the maximum iteration number, and is discarded. If the discarded source is denoted as x_i, then the scout bee looks for a new source of food to update x_i using the following process:

$$x_{i,j} = \delta_l^j + r(\delta_u^j - \delta_l^j) \tag{3}$$

where r is a number that is randomly generated using a normal distribution in the range [0, 1].

Generally, the BA algorithm excels at search space exploitation, however, it may become captivated in local optima, hindering its ability to conduct effective global searches. Since the BA algorithm relies solely on random walks, rapid convergence is not ensured. To address this limitation, we introduce a key enhancement in our proposed algorithm to augment the population's diversity and mitigate the risk of being trapped in local optima. This enhancement involves integrating the mutation operator from ABC, which accelerates convergence. By doing so, the BA-ABC algorithm becomes more suitable for a broader array of practical applications while retaining the favorable traits of the fundamental BA algorithm. Thus, the distinction between BA-ABC and BA algorithms lies in the utilization of the mutation operator to refine the BA algorithm by producing a new solution for each bat. Consequently, this methodology enables the exploration of novel search spaces through ABC's mutation mechanism while leveraging population information with the BA algorithm, thereby circumventing the pitfalls of local optima in the BA algorithm.

In the BA-ABC algorithm, all stages involve additional steps to initialize the population and exchange information within the swarm. Initially, key parameters such as the population size N, the solution space dimensions n, and the max number of iterations it_{max} as the termination criteria are defined. Subsequently, other parameters like f_i, v_i, A_i, and r_i are computed. The flowchart, given by Figure 7, consists of three phases: population initialization, BA phase, and ABC phase. The population initialization phase is executed once in start while the BA and ABC phases are executed

Algorithm 1 BA-ABC Algorithm

Input: $N, n, f_{min}, f_{max}, [\delta_u, \ldots, \delta_l], it_{max}, V_i, A_i, r_i, X_{ij}$ and random numbers $(a, b, r_1, r_2, \beta, \phi)$
Output: $\mathbf{X}_b$

Initialization:
- Generate initial bat population $\mathbf{X}_i$ using chaotic mapping:
  $\mathbf{X}(k+1) = \text{mod}\left(\mathbf{b} + \mathbf{X}(k) + -\frac{\mathbf{a}}{2\pi}\sin(2\pi\mathbf{X}(k)), \mathbf{1}\right)$
Evaluate fitness and select current best $\mathbf{X}^*$
while $it < it_{\max}$ **do**
  Adjust Frequency and Update Velocities:
    $\mathbf{f}_i = (\mathbf{f}_{\max} - \mathbf{f}_{\min}) \times \beta + \mathbf{f}_{\min}$
    $\mathbf{V}_i^{t+1} = \mathbf{V}_i^t + (\mathbf{X}_i^t - \mathbf{X}^*) \times \mathbf{f}_i$
  Update Positions/Solutions:
    $\mathbf{X}_i^{t+1} = \mathbf{X}_i^t + \mathbf{V}_i^{t+1}$
  if $r_1 > \mathbf{r}_i$ **then**
    - Move agents using updated f_i, v_i, and x_i
  end if
  Evaluate Fitness:
    Accept or reject $\mathbf{X}_i^{t+1}$
  if $r_2 < A_i$ & $f(\mathbf{X}_i) \leq f(\mathbf{X}^*)$ **then**
    - $\uparrow \mathbf{r}_i$ and $\downarrow A_i$
  end if
  Evaluate fitness:
    Select current best $\mathbf{X}^*$
  Calculate r using logistic chaotic mapping:
    $\mathbf{X}(j+1) = \mathbf{a} \times \mathbf{X}(j) \times (\mathbf{X}(j) - 1)$, while $\mathbf{X}(0) = \mathbf{X}^*$
  Select $\mathbf{X}_{ij}$ using r
  Mutate Search Agents:
    $\mathbf{V}_{ij} = \mathbf{X}_{ij} + \phi_{ij} \times (\mathbf{X}_{ij} - \mathbf{X}_{kj})$
  Apply Greedy Selection
  Find New Solution, Evaluate Fitness & select $\mathbf{X}^*$
end while
Return the Best Solution $\mathbf{X}_{gb} = \mathbf{X}^*$

repeatedly, in sequence. In the population initialization phase, chaotic mapping is employed to produce initial possible solutions from a broader search space, aiming to enhance the fitness of the initial population. The strategy for population initialization is elaborated in Algorithm 1.

The initial population X is generated using the logistic chaotic mapping within the bounds of the search space. The logistic chaotic function, utilized here is expressed as:

$$x(p+1) = x(p) * (p-1) * 4 \tag{4}$$

where, p is the iteration number, and the initial value x(0) is randomly chosen.

After generating the initial population, the algorithm follows the standard BA procedure, updating its parameters and the current positions of search agents using the steps 9, 10 and 12 given in the algorithm, and Equations 5–7. As mentioned earlier, both employed and onlooker bees in ABC share information about the possible solutions in the swarm and adjust previous solutions using Equation 1. To enhance

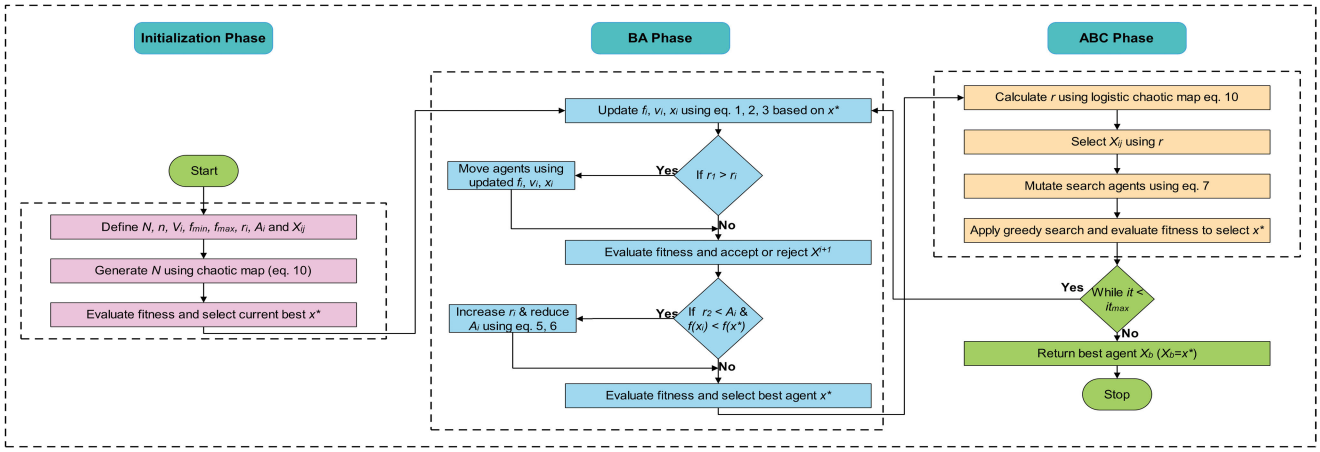


FIGURE 7. Flow of the introduced BA-ABC algorithm.

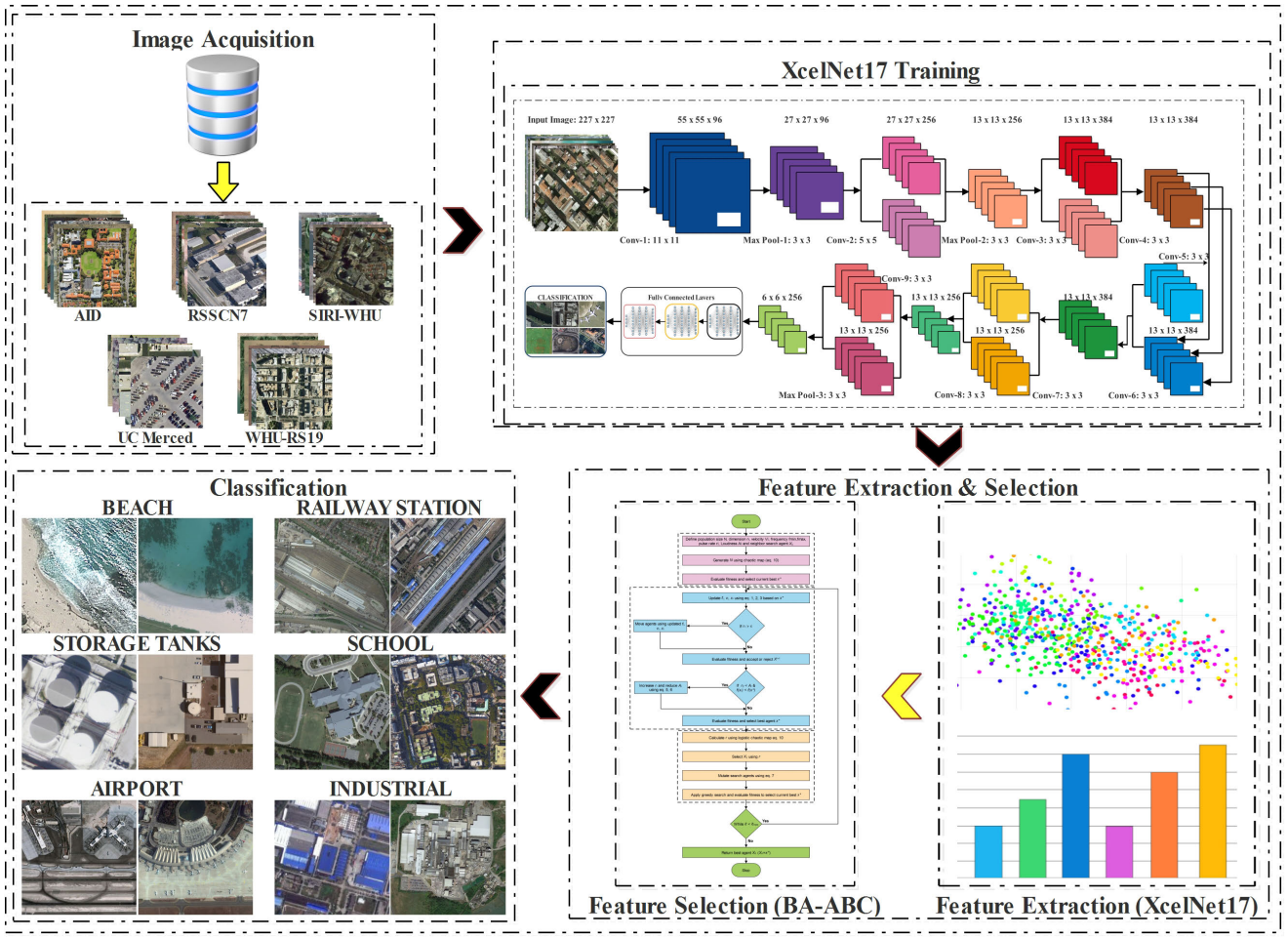


FIGURE 8. Work flow of the proposed framework.

random behavior and prevent repetition, logistic chaotic mapping defined in Equation 4, is utilized to determine the X_{ji} term in Equation 1. This approach provides enhanced exploitation possibilities by allowing for the selection of random neighboring solutions and positions for information exchange.

$$X_u = X_o + \epsilon A^t \tag{5}$$

here X_u is the new position, X_o is the previous position, ϵ is a randomly generated number in the [−1,1] range, and A^t denotes the bat's average loudness at time t.

If $f(X_i) < f(X^*)$ and the randomly generated number $r_2 < A_i$, then consider the new solution valid. Further update A_i and r_i, as below:

$$A_i^{t+1} = \alpha \times A_i^t \tag{6}$$

$$r_i^t = r_i^0[1-e^{-\gamma t}] \tag{7}$$

here A_i^{t+1} and A_i^t are loudness at times t+1 and t, respectively; r_i^0 and r_i^t are the pulse rates initially and at time t. As $t \to \infty$, $A_i^t \to 0$ and $r_i^t \to r_i^0$.

The entire exploration and exploitation process, comprising the BA, and ABC phases, iterates for a specified number it_{max}, after which the best solution is returned as a result. Through the search equation of ABC, i.e., Equation 1 in BA, the search capability is enhanced globally as each member of the swarm has the chance to exchange information with others. This fosters the maintenance of necessary exploration and exploitation, mitigates the issue of diversity, and prevents premature convergence. Moreover, it provides a greater chance to escape local optima and find the global best solution. Figure 8 presents the detailed workflow of the process utilizing the proposed XcelNet17 architecture and the BA-ABC algorithm.

V. RESULTS

In this section, we assess the performance of XcelNet17 and the BA-ABC algorithm for RS image classification. This assessment is accomplished in two phases through a series of experiments. In the first phase, we assess the XcelNet17 performance while BA-ABC is evaluated in the next. To ensure a fair comparison of execution times, the entire experimentation is conducted on a Windows-11 PC, equipped with Core i7 6700 (Intel processor) functioning at 3.4 GHz, 16GB RAM, and NVIDIA GeForce GTX 1050 Ti GPU for both training and testing processes. This implementation is performed on MATLAB R2022b.

A. PERFORMANCE METRICS

Several variables are used in the evaluation process, including Overall Accuracy (OA), F-1 Score, Recall, Specificity (Spe), and Precision. These metrics provide a comprehensive evaluation of the prediction performance of a model.

1) OVERALL ACCURACY (OA)

OA is an important metric when evaluating the effectiveness of classification models. It is especially useful in the context of image classification or other predictive modeling applications. It is calculated as the ratio of accurate predictions to the total number of predictions. Mathematically, it can be expressed as:

$$OA = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \tag{8}$$

where T_P represents correct positive predictions, T_N represents correct negative predictions, F_P represents incorrect negative predictions, and F_N represents incorrect positive predictions. The importance of Overall Accuracy lies in its clear explanation. It shows the percentage of correct predictions generated by the model out of all predictions. This makes it a straightforward way to determine a model's ability to predict outcomes.

2) F-1 SCORE

The F-1 Score is an important measure when assessing the performance of image classification models, particularly in situations with unevenly distributed data. It is the harmonic mean of precision and recall, two metrics that evaluate the accuracy and completeness of model predictions. Here's the formula for the F-1 Score:

$$F1 = 2 \times \frac{(Pr \times Re)}{Pr + Re} \tag{9}$$

where Pr (Precision) is the ratio of true positive predictions to total predicted positives, and Re (Recall), also called Sensitivity or True Positive Rate, is the ratio of true positive predictions to actual positives. The F-1 Score is significant because it balances recall and precision. It is useful in situations when you need a single measurement for model comparison or where the costs associated with wrong positives and incorrect negatives are comparable. Because the F-1 Score is unaffected by a large number of true negatives, it is a better solution than accuracy for problems with uneven class distribution.

3) PRECISION

It plays a crucial role in evaluating the performance of image classification models, especially when false positives could have a major impact. It evaluates how well the model predicts positive outcomes by calculating the accuracy of those predictions. Here's the formula for precision:

$$Precision = \frac{T_P}{T_P + F_P} \tag{10}$$

where T_P (True Positives) refer to correct positive predictions, while F_P (False Positives) are incorrect positive predictions. The Precision is significant since it emphasizes how well the model predicts the positive class. A high degree of accuracy indicates that there are few false alarms and that the model can be relied upon to forecast a positive result. This is particularly important in scenarios where false positives might have serious repercussions, such as spam detection or medical diagnosis.

4) RECALL

Recall, also referred to as Sensitivity or True Positive Rate, is a key measurement in assessing the effectiveness of image classification models. It calculates the percentage of true positive cases correctly identified by the model. Here's the

TABLE 4. Performance evaluation of XcelNet17 on the AID.

Data Set	Network	Classifier	OA	F-1	Recall	Spe	Prescision	Acc
AID	AlexNet	SVM	96.100	0.960	0.961	0.997	0.961	0.961
		Med NN	94.900	0.948	0.949	0.998	0.948	0.949
		KNN	91.500	0.914	0.912	0.997	0.922	0.915
	VGG16	SVM	94.100	0.939	0.939	0.998	0.940	0.941
		Med NN	93.900	0.937	0.937	0.998	0.937	0.939
		KNN	92.200	0.922	0.918	0.997	0.922	0.922
	VGG19	SVM	93.200	0.930	0.932	0.998	0.930	0.932
		Med NN	92.200	0.919	0.920	0.997	0.919	0.922
		KNN	91.600	0.913	0.913	0.997	0.917	0.917
	ResNet-50	SVM	93.800	0.938	0.939	0.998	0.939	0.938
		Med NN	92.700	0.925	0.925	0.997	0.926	0.927
		KNN	93.000	0.928	0.928	0.998	0.930	0.930
	DarkNet-19	SVM	91.300	0.911	0.909	0.997	0.913	0.913
		Med NN	91.300	0.910	0.911	0.997	0.910	0.913
		KNN	91.800	0.916	0.915	0.997	0.919	0.919
	XcelNet17	SVM	96.300	0.963	0.963	0.999	0.963	0.963
		Med NN	95.200	0.952	0.952	0.998	0.952	0.952
		KNN	94.700	0.947	0.947	0.998	0.947	0.947

TABLE 5. Performance evaluation of XcelNet17 on the RSSCN7.

Data Set	Network	Classifier	OA	F-1	Recall	Spe	Prescision	Acc
RSSCN7	AlexNet	SVM	89.900	0.899	0.899	0.983	0.899	0.899
		Med NN	89.000	0.890	0.890	0.982	0.890	0.890
		KNN	86.200	0.860	0.862	0.977	0.862	0.862
	VGG16	SVM	93.000	0.930	0.930	0.988	0.930	0.930
		Med NN	91.200	0.912	0.913	0.985	0.913	0.913
		KNN	92.100	0.921	0.921	0.987	0.921	0.921
	VGG19	SVM	90.700	0.907	0.907	0.985	0.907	0.907
		Med NN	91.800	0.918	0.918	0.986	0.918	0.918
		KNN	90.400	0.904	0.904	0.984	0.904	0.904
	ResNet-50	SVM	93.800	0.937	0.938	0.990	0.937	0.938
		Med NN	92.900	0.928	0.929	0.988	0.929	0.929
		KNN	93.200	0.932	0.932	0.989	0.932	0.932
	DarkNet-19	SVM	76.100	0.806	0.806	0.968	0.811	0.806
		Med NN	77.700	0.777	0.777	0.963	0.778	0.777
		KNN	75.700	0.758	0.757	0.959	0.760	0.757
	XcelNet17	SVM	96.400	0.964	0.964	0.994	0.964	0.964
		Med NN	95.400	0.954	0.954	0.992	0.954	0.954
		KNN	94.600	0.946	0.946	0.991	0.946	0.946

formula for recall:

$$Recall = \frac{T_P}{T_P + F_N} \tag{11}$$

where T_P refers to correctly identified positive cases and F_N refers to positive cases incorrectly classified as negative by the model. The importance of recall is centered on how well the model can identify all given examples. A high recall indicates that the model is efficient in identifying the positives and does not overlook many genuine positive instances. This is especially crucial in scenarios where not detecting positives could result in significant consequences.

TABLE 6. Performance evaluation of XcelNet17 on the SIRI-WHU.

Data Set	Network	Classifier	OA	F-1	Recall	Spe	Prescision	Acc
SIRI-WHU	AlexNet	SVM	90.800	0.908	0.908	0.992	0.908	0.908
		Med NN	87.300	0.872	0.873	0.988	0.873	0.873
		KNN	87.500	0.874	0.875	0.989	0.875	0.875
	VGG16	SVM	92.100	0.921	0.921	0.993	0.921	0.921
		Med NN	92.300	0.922	0.923	0.993	0.923	0.923
		KNN	92.700	0.927	0.927	0.993	0.927	0.927
	VGG19	SVM	94.400	0.944	0.944	0.995	0.944	0.944
		Med NN	95.400	0.954	0.954	0.996	0.954	0.954
		KNN	94.400	0.943	0.944	0.995	0.944	0.944
	ResNet-50	SVM	93.800	0.938	0.938	0.994	0.938	0.938
		Med NN	94.600	0.946	0.946	0.995	0.946	0.946
		KNN	95.000	0.950	0.950	0.995	0.950	0.950
	DarkNet-19	SVM	83.600	0.836	0.836	0.985	0.836	0.836
		Med NN	82.600	0.825	0.826	0.984	0.826	0.826
		KNN	80.300	0.803	0.803	0.982	0.803	0.803
	XcelNet17	SVM	97.100	0.971	0.971	0.997	0.971	0.971
		Med NN	96.900	0.969	0.969	0.997	0.969	0.969
		KNN	97.600	0.976	0.976	0.998	0.976	0.976

TABLE 7. Performance evaluation of XcelNet17 on the UC Merced.

Data Set	Network	Classifier	OA (Val)	F-1	Recall	Spe	Prescision	Acc
UC Merced	AlexNet	SVM	94.570	94.590	95.010	99.780	94.570	94.570
		Med NN	92.690	92.750	93.380	99.700	92.690	92.690
		KNN	93.400	93.350	93.870	99.730	93.400	93.400
	VGG16	SVM	94.570	94.600	95.010	99.780	94.570	94.570
		Med NN	94.800	94.850	95.480	99.770	94.800	94.800
		KNN	94.350	94.320	95.000	99.780	94.350	94.350
	VGG19	SVM	94.810	94.860	95.210	99.740	94.760	94.760
		Med NN	93.640	93.770	94.350	99.750	93.640	93.640
		KNN	95.040	95.120	95.420	99.790	95.040	95.040
	ResNet-50	SVM	93.860	93.790	94.030	99.690	93.810	93.810
		Med NN	92.450	92.400	92.980	99.690	92.450	92.450
		KNN	93.630	93.550	93.980	99.740	93.630	93.630
	DarkNet-19	SVM	89.810	89.900	90.520	99.490	89.760	89.760
		Med NN	89.120	89.370	90.320	99.520	89.120	89.120
		KNN	88.140	88.110	89.090	99.450	88.140	88.140
	Proposed	SVM	99.750	98.480	98.900	99.960	98.600	99.750
		Med NN	98.470	97.190	97.140	99.920	97.210	98.470
		KNN	97.690	96.430	97.140	99.900	96.030	97.690

For example, in medical assessments, overlooking an illness could have more severe impacts than incorrectly identifying one.

5) SPECIFICITY

Specificity, which is also referred to as the True Negative Rate, is a measurement used to assess how well an image classification model performs. It assesses the accuracy of identifying actual negative outcomes by the model. Specificity is calculated by:

$$Specificity = \frac{T_N}{T_N + F_P} \tag{12}$$

where T_N represents correctly identified negative instances and F_P represents negative instances mistakenly identified as positive. The specificity is a very important metric as it prioritizes the model's accuracy in detecting negative

TABLE 8. Performance evaluation of XcelNet17 on the WHU-RS19.

Data Set	Network	Classifier	OA	F-1	Recall	Spe	Prescision	Acc
WHU-RS19	AlexNet	SVM	95.595	95.509	95.662	99.804	95.696	95.595
		Med NN	94.619	94.666	95.325	99.764	94.706	94.619
		KNN	93.639	93.741	94.611	99.720	93.711	93.639
	VGG16	SVM	95.595	95.673	96.201	99.804	95.696	95.595
		Med NN	96.570	96.687	97.148	99.844	96.638	96.570
		KNN	95.610	95.728	96.419	99.819	95.663	95.610
	VGG19	SVM	92.624	92.712	93.506	99.640	92.666	92.624
		Med NN	92.609	92.615	94.042	99.627	92.746	92.609
		KNN	93.589	93.549	94.075	99.670	93.566	93.589
	ResNet-50	SVM	80.743	82.652	89.681	98.989	80.608	80.743
		Med NN	95.580	95.566	95.974	99.789	95.537	95.580
		KNN	94.619	94.593	94.921	99.764	94.610	94.619
	DarkNet-19	SVM	96.585	96.571	96.772	99.858	96.645	96.585
		Med NN	96.095	96.088	96.488	99.836	96.084	96.095
		KNN	95.125	95.251	95.945	99.801	95.051	95.125
	XcelNet17	SVM	99.900	98.245	98.734	99.890	98.296	99.900
		Med NN	98.665	96.886	97.258	99.850	96.907	98.665
		KNN	99.155	98.249	98.709	99.873	97.769	99.155

TABLE 9. Results of ANOVA with classifiers on AID for XcelNet17.

Source of Variance	SS	df	MSE	F-statistic	p-value
Inter-groups	5.001	2	2.500	0.639	0.559
Intra-group	23.451	6	3.908		
Total	28.452	8			

TABLE 10. Results of ANOVA with classifiers on RSSCN7 for XcelNet17.

Source of Variance	SS	df	MSE	F-statistic	p-value
Inter-groups	3.369	2	1.684	0.189	0.833
Intra-group	53.547	6	8.925		
Total	56.916	8			

TABLE 11. Results of ANOVA with classifiers on SIRI-WHU for XcelNet17.

Source of Variance	SS	df	MSE	F-statistic	p-value
Inter-groups	0.837	2	0.418	0.088	0.917
Intra-group	28.608	6	4.768		
Total	29.445	8			

TABLE 12. Results of ANOVA with classifiers on UC Merced for XcelNet17.

Source of Variance	SS	df	MSE	F-statistic	p-value
Inter-groups	3.128	2	1.564	0.403	0.685
Intra-group	23.257	6	3.876		
Total	26.385	8			

TABLE 13. Results of ANOVA with classifiers on WHU-RS19 for XcelNet17.

Source of Variance	SS	df	MSE	F-statistic	p-value
Inter-groups	0.040	2	0.020	0.0057	0.994
Intra-group	21.102	6	3.517		
Total	21.142	8			

instances. This is especially significant in situations where false negatives are less expensive than false positives. In a security system utilizing facial recognition for access control, high specificity ensures that unauthorized persons are seldom mistaken for authorized ones, thus upholding the system's integrity.

B. XCELNET17 PERFORMANCE EVALUATION

In the first phase of experiments, we ascertain the efficiency and accuracy of the XcelNet17, against some well-established networks including AlexNet [17], VGG16 [25], VGG19 [25], ResNet50 [18], and DarkNet19 [26]. We have utilized five datasets including AID, RSSCN7, SIRI-WHU, UC Merced, and WHU RS-19 as mentioned in Sect. IV-A. During the experiments, we utilize a mini-batch size of 16 and employed the Stochastic Gradient Descent with Momentum (SGDM) as the optimizer. We apply a decay (drop) factor of 0.5 and a drop period of 7 parameters while maintaining a learning rate of e^{-4}. Each model undergoes training for 500 epochs, with the dataset randomly divided into 20% for testing and 80% for training. It is clarified that this refers to the initial dataset splitting which is used for model training. Once the features were extracted, we utilized 10-fold cross-validation to train the classifiers in MATLAB, which is the actual step in training-testing. The reported accuracies in the manuscript refer to the output yielded after this whole process.

Tables 4, 5, 6, 7, and 8 show results for the five selected datasets. The list of classifiers is shown against the metric chosen for evaluations.

These results show how various combinations of network models and classifiers work on the datasets. The evaluation

TABLE 14. Results of Bonferroni test with classifiers on AID for XcelNet17.

Group A	Group B	Lower Limit	Difference	Upper Limit	P-value
SVM-Cubic	Medium NN	-3.886	1.067	6.019	0.793
SVM-Cubic	KNN-Cubic	-3.136	1.817	6.769	0.535
Medium NN	KNN-Cubic	-4.203	0.750	5.703	0.890

TABLE 15. Results of Bonferroni test with classifiers on RSSCN7 for XcelNet17.

Group A	Group B	Lower Limit	Difference	Upper Limit	P-value
SVM-Cubic	Medium NN	-6.484	1.000	8.484	0.913
SVM-Cubic	KNN-Cubic	-6.017	1.467	8.951	0.825
Medium NN	KNN-Cubic	-7.017	0.467	7.951	0.980

TABLE 16. Results of Bonferroni test with classifiers on SIRI-WHU for XcelNet17.

Group A	Group B	Lower Limit	Difference	Upper Limit	P-value
SVM-Cubic	Medium NN	-5.202	0.268	5.738	0.988
SVM-Cubic	KNN-Cubic	-5.940	-0.470	5.000	0.963
Medium NN	KNN-Cubic	-6.208	-0.738	4.732	0.911

TABLE 17. Results of Bonferroni test with classifiers on UCM for XcelNet17.

Group A	Group B	Lower Limit	Difference	Upper Limit	P-value
SVM-Cubic	Medium NN	-3.990	0.943	5.875	0.832
SVM-Cubic	KNN-Cubic	-3.514	1.419	6.351	0.670
Medium NN	KNN-Cubic	-4.456	0.476	5.408	0.953

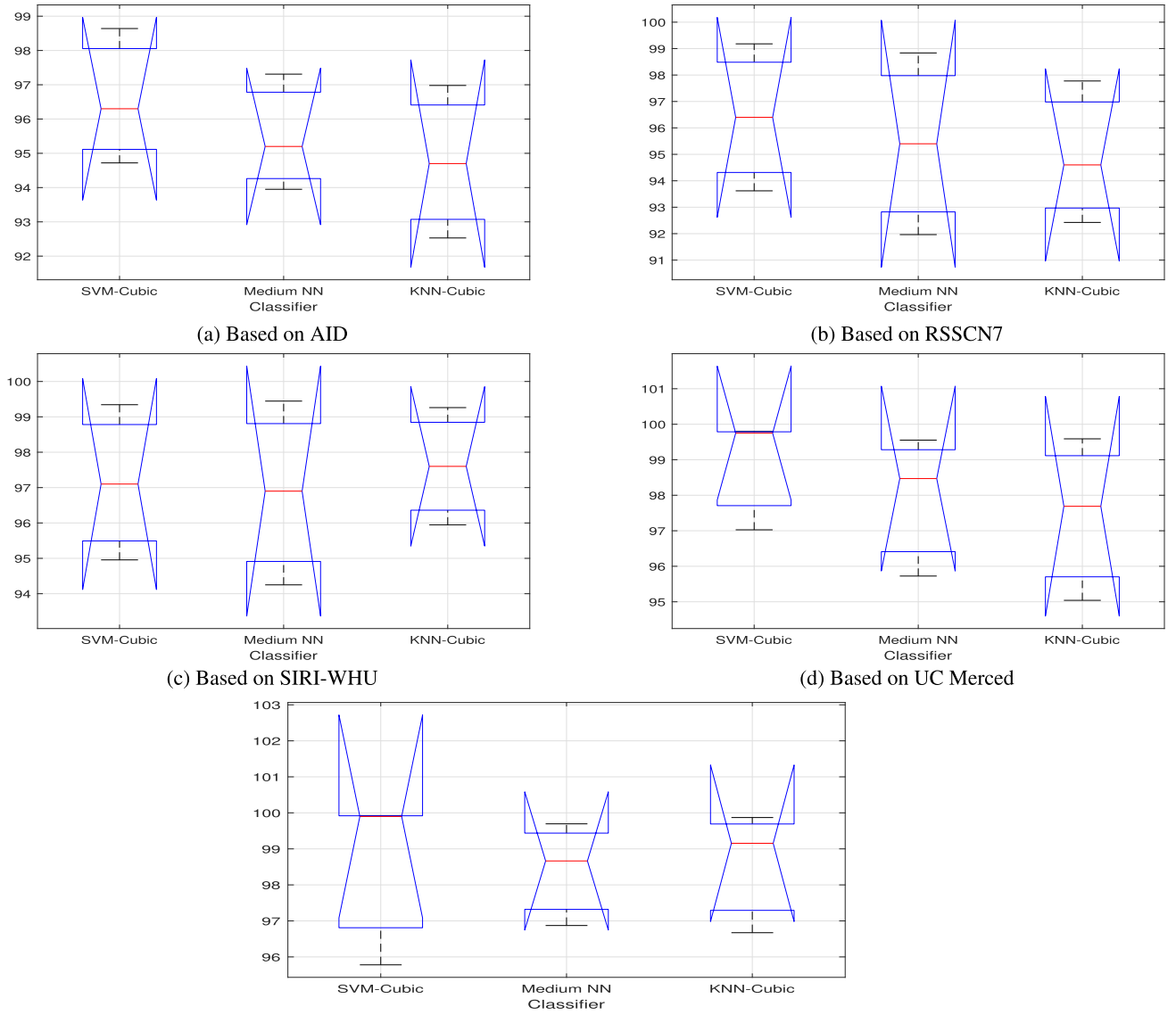
TABLE 18. Results of Bonferroni test with classifiers on WHU-RS19 for XcelNet17.

Group A	Group B	Lower Limit	Difference	Upper Limit	P-value
SVM-Cubic	Medium NN	-4.574	0.124	4.823	0.996
SVM-Cubic	KNN-Cubic	-4.729	-0.031	4.668	1.000
Medium NN	KNN-Cubic	-4.853	-0.155	4.543	0.994

process makes use of several factors, including Overall Accuracy (OA), F-1 Score, Recall, Specificity (Spe) and Precision. These indicators help in a thorough assessment of a model's predictive performance.

From the results, we can see that the SVM performs best, when compared to other classifiers. When handling the features extracted by deep learning architectures for classification, SVM has proven to be most robust and effective. Among the five datasets, XcelNet17 emerges as the superior neural network architecture especially when paired with the SVM classifier. This shows its ability to achieve high precision in similar tasks. All the models have consistently high specificity scores, demonstrating how well they can identify negative cases. Furthermore, the balanced distribution of F-1, Recall, and Precision scores highlights their overall effectiveness in preserving accuracy and precisely recognizing true positives.

Tables 4 through Table 7 demonstrate that XcelNet17 performs exceptionally well on all datasets especially when paired with the SVM classifier, as mentioend earlier.



(a) Based on AID

(b) Based on RSSCN7

(c) Based on SIRI-WHU

(d) Based on UC Merced

(e) Based on WHU-RS19

FIGURE 9. Accuracy box-plots of different Classifiers for XcelNet17.

The combination manages to achieve an OA rate of 94.6% ∼ 99.9% over all configurations tested. This highlights XcelNet17's remarkable efficiency as a FS method.

1) STATISTICAL ANALYSIS OF XCELNET17 RESULTS

The essence of conducting statistical analysis is to provide a degree of assurance in the selected network and to determine the general validity of the obtained results. To achieve this, we employ the analysis of variance (ANOVA) [63] method with Bonferroni post-hoc testing [64]. This approach allows us to assess whether the selected framework demonstrates significant improvement. ANOVA performs a comparison of the means of multiple distributions by evaluating variances with-in and between the groups.

Before proceeding with ANOVA, it's essential to verify the equality and normality of variance assumptions. The Shapiro-Wilk test [65] is employed to accomplish the assumption of normality and the Bartlett's test [65] is employed to determine whether variances are homogeneous. The value of 0.01 is used for the significance level (α).

Let $\overline{x_1}$, $\overline{x_2}$, and $\overline{x_3}$ represent the means of XcelNet17 (OA with three different classifiers). $\overline{x_1} = \overline{x_2} = \overline{x_3}$, is the null hypothesis (H_0), while $x_i \neq x_j$, is the alternative hypothesis (H_a), where (i, j) belongs to 1, 2, 3. If the calculated p-value is lesser than α, i.e., $p < \alpha$, the null hypothesis is rejected. This indicates significant differences among the means. In such cases, the Bonferroni posthoc technique (also known as the multiple comparison technique) is applied to validate all pairwise means and find which ones are significantly apart.

TABLE 19. Accuracy Comparison of BA-ABC with 5 cutting edge algorithms.

Data Set	Classifier	WOA	GWO	BA	ABC	ACO	BA-ABC
WHU-RS19	SVM	82.500	87.500	82.500	80.000	82.500	**95.000**
	KNN	77.500	90.000	92.500	87.500	77.500	**97.500**
	Med NN	87.500	87.500	90.000	90.000	77.500	**97.500**
UC Merced	SVM	96.100	94.600	96.100	97.900	96.400	**98.200**
	KNN	95.500	95.800	97.300	**97.900**	95.500	**97.900**
	Med NN	94.000	**96.700**	94.900	96.400	95.500	**96.700**
RSSCN7	SVM	97.300	96.700	97.800	96.200	96.700	**98.700**
	KNN	97.100	96.900	96.700	96.900	95.100	**98.700**
	Med NN	97.300	96.900	97.300	96.900	96.400	**98.900**
SIRI-WHU	SVM	97.900	97.700	98.200	97.900	98.200	**99.000**
	KNN	97.900	97.900	97.500	97.700	94.500	**99.000**
	Med NN	96.600	**97.700**	97.300	96.900	97.400	**97.700**
AID	SVM	84.000	83.000	87.200	85.200	88.200	**88.300**
	KNN	82.500	82.800	87.800	85.200	86.200	**88.500**
	Med NN	81.500	79.800	82.800	80.200	82.500	**85.200**

TABLE 20. F1 Score Comparison of BA-ABC with 5 cutting edge algorithms.

Data Set	Classifier	WOA	GWO	BA	ABC	ACO	BA-ABC
WHU-RS19	SVM	0.800	0.875	0.700	0.900	0.925	0.935
	KNN	0.750	0.900	0.875	0.925	0.800	0.941
	Med NN	0.950	0.900	0.925	0.900	0.900	0.950
UC Merced	SVM	0.964	0.962	0.967	0.935	0.958	0.971
	KNN	0.970	0.958	0.961	0.944	0.955	0.977
	Med NN	0.947	0.963	0.935	0.952	0.979	0.985
RSSCN7	SVM	0.978	0.982	0.982	0.984	0.978	0.986
	KNN	0.969	0.984	0.975	0.973	0.975	0.985
	Med NN	0.984	0.963	0.969	0.971	0.969	0.985
SIRI-WHU	SVM	0.984	0.987	0.981	0.974	0.974	0.989
	KNN	0.966	0.982	0.984	0.975	0.964	0.985
	Med NN	0.958	0.966	0.959	0.974	0.976	0.995
AID	SVM	0.853	0.882	0.861	0.820	0.870	0.885
	KNN	0.810	0.848	0.861	0.829	0.855	0.870
	Med NN	0.771	0.812	0.814	0.782	0.791	0.820

In the first set of experiments, the three classifiers (SVM, Medium NN, KNN) are employed separately for each of the five datasets. The Shapiro-Wilk test yields p-values of 0.787, 0.720, and 0.973 for AID; 1.0, 1.0, and 0.795 for RSSCN7; 0.975, 0.978, and 0.997 for SIRI-WHU; 0.030, 0.531, and 0.819 for UC Merced, and 0.012, 0.704, and 0.410 for WHU-RS19 datasets. All of these values are above the α value. We are unable to reject the null hypothesis (H_0) on the basis of these results of equality and normality of variances and come to the conclusion that we have data that has normal distribution and has a variance of homogeneous kind.

In Tables 9, 10, 11, 12, and 13 statistical results are provided for the five selected datasets using the ANOVA test.

TABLE 21. Comparison of BA-ABC with 5 state-of-the-art methods on Sensitivity.

Data Set	Classifier	WOA	GWO	BA	ABC	ACO	BA-ABC
WHU-RS19	SVM	0.800	0.875	0.700	0.900	0.925	0.935
	KNN	0.750	0.900	0.875	0.925	0.800	0.941
	Med NN	0.950	0.900	0.925	0.900	0.900	0.950
UC Merced	SVM	0.964	0.964	0.976	0.935	0.958	0.979
	KNN	0.970	0.958	0.962	0.949	0.955	0.982
	Med NN	0.947	0.963	0.935	0.952	0.979	0.982
RSSCN7	SVM	0.978	0.982	0.982	0.984	0.978	0.986
	KNN	0.970	0.984	0.975	0.973	0.975	0.986
	Med NN	0.984	0.964	0.969	0.971	0.969	0.985
SIRI-WHU	SVM	0.984	0.986	0.981	0.974	0.973	0.988
	KNN	0.966	0.982	0.984	0.977	0.964	0.986
	Med NN	0.958	0.966	0.960	0.973	0.977	0.995
AID	SVM	0.872	0.885	0.865	0.823	0.870	0.886
	KNN	0.842	0.848	0.865	0.830	0.855	0.870
	Med NN	0.783	0.823	0.815	0.786	0.789	0.825

TABLE 22. Comparison of BA-ABC with 5 state-of-the-art methods on Specificity.

Data Set	Classifier	WOA	GWO	BA	ABC	ACO	BA-ABC
WHU-RS19	SVM	0.989	0.993	0.983	0.994	0.996	0.996
	KNN	0.986	0.994	0.993	0.996	0.989	0.997
	Med NN	0.997	0.994	0.996	0.994	0.994	0.997
UC Merced	SVM	0.998	0.998	0.999	0.997	0.998	0.999
	KNN	0.999	0.998	0.998	0.997	0.998	0.999
	Med NN	0.997	0.998	0.997	0.998	0.999	0.999
RSSCN7	SVM	0.996	0.997	0.997	0.997	0.996	0.997
	KNN	0.995	0.997	0.996	0.996	0.996	0.997
	Med NN	0.997	0.994	0.995	0.995	0.995	0.998
SIRI-WHU	SVM	0.999	0.999	0.998	0.998	0.998	0.999
	KNN	0.997	0.998	0.999	0.998	0.997	0.999
	Med NN	0.996	0.997	0.996	0.998	0.998	1.000
AID	SVM	0.995	0.996	0.995	0.994	0.996	0.996
	KNN	0.993	0.995	0.995	0.994	0.995	0.996
	Med NN	0.992	0.994	0.994	0.993	0.993	0.995

The five different parameters considered are sum of squared deviation (SS), mean squared error (MSE), degree of freedom (df), p-value, and F statistic.

SS measures the total variation, variation within groups, and variation between groups. df reflects the number of independent values that can vary in the calculation. For inter-groups df is 2, suggesting three groups are being compared (since df = number of groups - 1). Within-groups df is 6, which implies there are a total of 7 observations spread across the groups. The total df is the sum of inter-groups and within-groups df, here totaling 8.

MSE is calculated by dividing the SS by the corresponding df. It represents the average variance within groups (2.50) and the average variance between groups (3.908). The F-statistic,

TABLE 23. Comparison of BA-ABC with 5 state-of-the-art methods on Precision.

Data Set	Classifier	WOA	GWO	BA	ABC	ACO	BA-ABC
WHU-RS19	SVM	0.800	0.875	0.700	0.900	0.925	0.935
	KNN	0.750	0.900	0.875	0.925	0.800	0.941
	Med NN	0.950	0.900	0.925	0.900	0.900	0.950
UC Merced	SVM	0.964	0.962	0.965	0.935	0.956	0.971
	KNN	0.970	0.958	0.961	0.943	0.955	0.977
	Med NN	0.949	0.964	0.935	0.952	0.979	0.985
RSSCN7	SVM	0.978	0.982	0.982	0.984	0.978	0.986
	KNN	0.969	0.984	0.975	0.973	0.975	0.985
	Med NN	0.984	0.962	0.968	0.971	0.968	0.987
SIRI-WHU	SVM	0.984	0.987	0.981	0.974	0.974	0.989
	KNN	0.966	0.982	0.984	0.975	0.964	0.985
	Med NN	0.958	0.966	0.959	0.974	0.974	0.995
AID	SVM	0.856	0.886	0.859	0.815	0.870	0.888
	KNN	0.802	0.848	0.859	0.830	0.855	0.870
	Med NN	0.776	0.815	0.815	0.790	0.803	0.820

TABLE 24. ANOVA results on classifiers with AID.

Source of Variance	SS	df	MSE	F-statistic	p-value
Inter-groups	11.390	2	5.695	0.704	0.531
Intra-group	48.533	6	8.089		
Total	59.923	8			

TABLE 25. ANOVA results on classifiers with RSSCN7.

Source of Variance	SS	df	MSE	F-statistic	p-value
Inter-groups	0.249	2	0.125	0.019	0.981
Intra-group	38.780	6	6.463		
Total	39.029	8			

TABLE 26. ANOVA results on classifiers with SIRI-WHU.

Source of Variance	SS	df	MSE	F-statistic	p-value
Inter-groups	2.227	2	1.114	0.121	0.888
Intra-group	55.105	6	9.184		
Total	57.333	8			

TABLE 27. ANOVA results on classifiers with UC Merced.

Source of Variance	SS	df	MSE	F-statistic	p-value
Inter-groups	3.363	2	1.681	0.235	0.798
Intra-group	43.002	6	7.167		
Total	46.365	8			

TABLE 28. ANOVA results on classifiers with WHU-RS19.

Source of Variance	SS	df	MSE	F-statistic	p-value
Inter-groups	5.978	2	2.989	0.388	0.694
Intra-group	46.186	6	7.698		
Total	52.164	8			

which measures whether group means differ significantly from one another, is a ratio of the variation among groups to the variance within a group. An F-statistic of 0.559 suggests that the inter-group variance is not much larger than the intra-group variance. Similar results/interpretations are obtained with other datasets as well.

If the null hypothesis is correct, the p-value indicates the likelihood of seeing the F-statistic (i.e., there are no differences between group means). The obtained p-values of 0.559, 0.833, 0.917, 0.685, and 0.994 are well above the typical significance threshold of 0.01, demonstrating that the null hypothesis cannot be rejected due to a lack of statistical evidence. Further, the Bonferroni posthoc is performed to validate these results. The Bonferroni post-hoc test is a method used to determine significant differences between multiple groups in ANOVA. This test involves conducting pairwise comparisons between all possible pairs of group means to identify specific pairs that differ significantly from each other.

The tables 9, 10, 11, 12, and 13 make it evident that the p-value exceeds the specified α. Thus the ANOVA and the Bonferroni posthoc test results have confirmed $p > \alpha$. Based on the pairwise comparison's results outlined in Tables 14, 15, 16, 17, and 18, we can assert with 95% confidence that the mean accuracies are within a close bound. Thus, the proposed DL architecture (XcelNet17) has consistent performance across all the datasets despite the chosen classifier. Therefore, our proposed approach, utilizing XcelNet17 with the three classifiers, exhibits significantly superior performance compared to conventional methods.

TABLE 29. Results of Bonferroni test with classifiers on AID for BA-ABC.

Group A	Group B	Lower Limit	Difference	Upper Limit	P-value
SVM-Cubic	KNN-Cubic	-6.868	0.257	7.382	0.993
SVM-Cubic	Medium NN	-4.620	2.505	9.630	0.560
KNN-Cubic	Medium NN	-4.878	2.247	9.372	0.622

TABLE 30. Results of Bonferroni test with classifiers on RSSCN7 for BA-ABC.

Group A	Group B	Lower Limit	Difference	Upper Limit	P-value
SVM-Cubic	KNN-Cubic	-6.404	-0.035	6.334	1.000
SVM-Cubic	Medium NN	-6.738	-0.369	6.000	0.983
KNN-Cubic	Medium NN	-6.703	-0.334	6.035	0.986

TABLE 31. Bonferroni post-hoc test results for three classifiers on SIRI-WHU for BA-ABC.

Group A	Group B	Lower Limit	Difference	Upper Limit	P-value
SVM-Cubic	KNN-Cubic	-7.815	-0.223	7.369	0.996
SVM-Cubic	Medium NN	-6.666	0.926	8.518	0.927
KNN-Cubic	Medium NN	-6.443	1.149	8.741	0.890

TABLE 32. Bonferroni post-hoc test results for three classifiers on UC Merced for BA-ABC.

Group A	Group B	Lower Limit	Difference	Upper Limit	P-value
SVM-Cubic	KNN-Cubic	-6.403	0.304	7.011	0.989
SVM-Cubic	Medium NN	-5.285	1.422	8.129	0.799
KNN-Cubic	Medium NN	-5.589	1.118	7.825	0.869

TABLE 33. Bonferroni post-hoc test results for three classifiers on WHU-RS19 for BA-ABC.

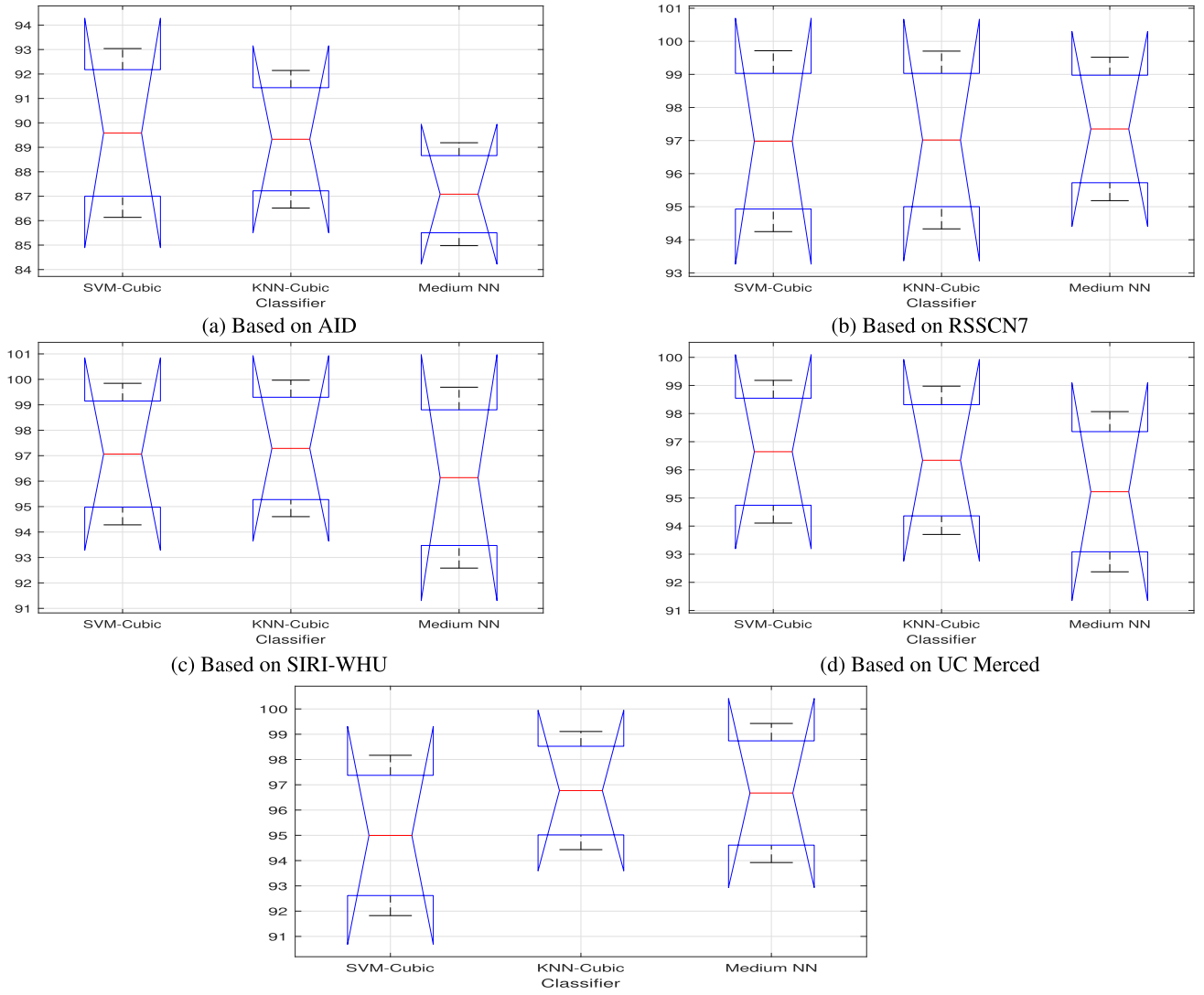
Group A	Group B	Lower Limit	Difference	Upper Limit	P-value
SVM-Cubic	KNN-Cubic	-8.726	-1.775	5.176	0.726
SVM-Cubic	Medium NN	-8.629	-1.679	5.272	0.750
KNN-Cubic	Medium NN	-6.854	0.096	7.047	0.999

Additionally, Figure 9 illustrates the effectiveness of these three classifiers across the five datasets.

C. BA-ABC PERFORMANCE EVALUATION

The performance of the BA-ABC algorithm is evaluated in this section. We have used the XcelNet17 CNN architecture as the feature extractor and compared BA-ABC performance on feature selection problems with five other population-based optimization methods including WOA, GWO, ACO, BA, and ABC as already mentioned in Sect. II.

The experiments are conducted on the same five most popular RS image classification datasets i.e. AID, WHU-RS19, UC Merced, RSSCN7, and SIRI-WHU as mentioned in Sect. IV-A. A comparison of the BA-ABC algorithm with five cutting-edge techniques is presented in Table 19 on accuracy. It is evident that the BA-ABC has managed to achieve higher accuracy than the contemporary algorithms. In some cases, however, the accuracy achieved by multiple are equal; they are highlighted in bold.



(a) Based on AID

(b) Based on RSSCN7

(c) Based on SIRI-WHU

(d) Based on UC Merced

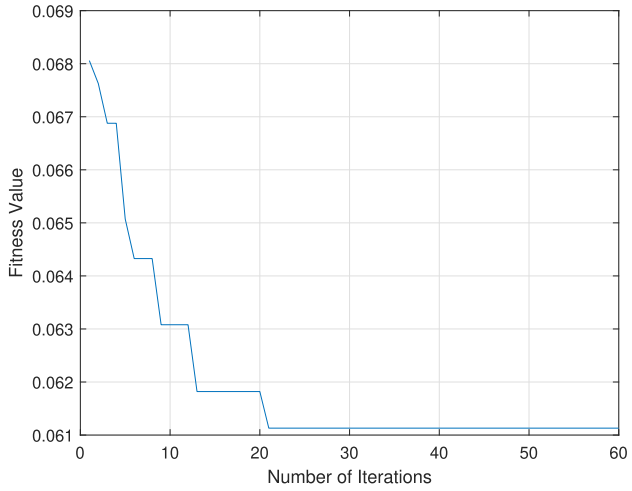
(e) Based on WHU-RS19

FIGURE 10. Accuracy box-plots of used classifiers for BA-ABC.

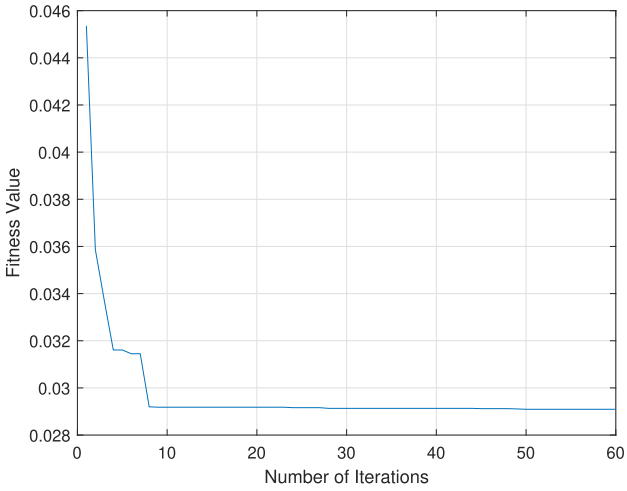
Similarly, Tables 20, 21, 22, and 23 show the comparison with five methods on five datasets utilizing F1 score, sensitivity, specificity, and precision, respectively. These results showcase the effectiveness of BA-ABC algorithm.

The accuracy results depict the performance of various classifiers (SVM-Cubic, KNN-Cubic, and Medium Neural Network) on different datasets. Across various datasets, SVM consistently demonstrates robust performance, achieving accuracies ranging from 82.5% to 99.0%. Notably, on the WHU-RS19 dataset, SVM attains accuracies between 82.5% and 95.0%, showcasing its effectiveness across diverse datasets. KNN performs competitively as well, with accuracy varying between 77.5% and 99.0%. Its remarkable accuracy of 99% on the SIRI-WHU dataset demonstrates its ability to precise classification. The accuracy of the Medium Neural Network varies greatly between datasets, despite its typically good performance. The results highlight the impact of classifier choice on classification accuracy. SVM and KNN have proved to be strong performers on different datasets. On the other hand, the Medium Neural Network's performance seems to be more dependent on the features of the dataset. The comparison of F1 score is displayed in Table 20 below:

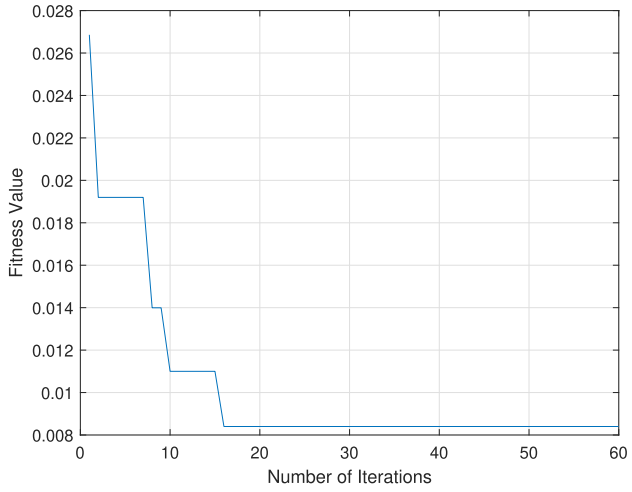
The F-1 score is another useful indicator for evaluating the effectiveness of classification models across different datasets and classifiers. The SVM has constantly high F-1 scores, indicating a well-balanced trade-off between precision and recall. SVM performs remarkably well on datasets such as SIRI-WHU and RSSCN7. Here, it continuously obtains F-1 values higher than 0.97, demonstrating its reliability for precise classification. Furthermore, both the KNN and the Medium Neural Network demonstrate



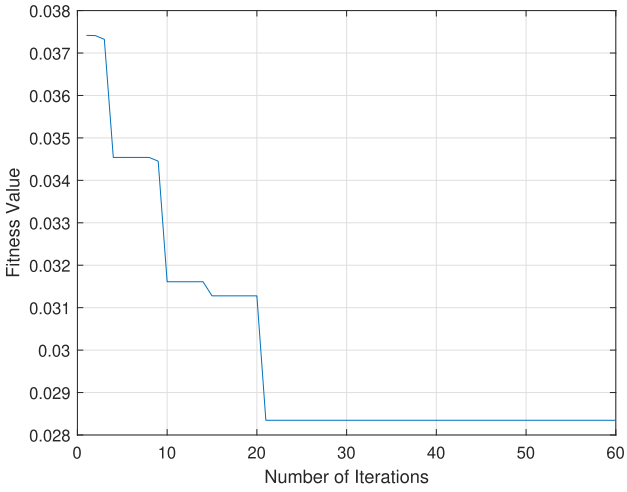
(a) Based on AID



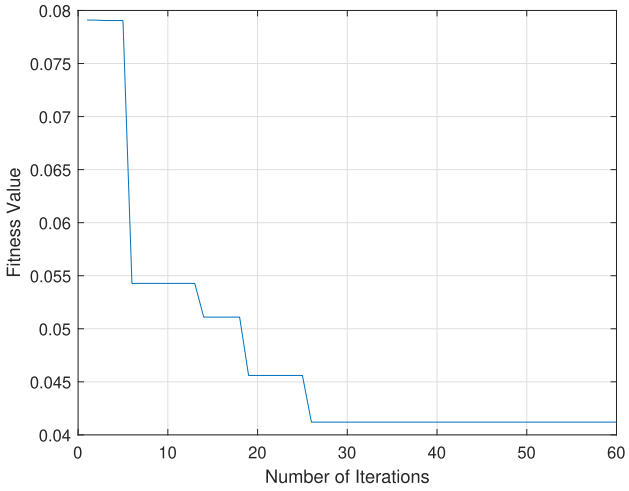
(b) Based on RSSCN7



(c) Based on SIRI-WHU



(d) Based on UC Merced



(e) Based on WHU-RS19

FIGURE 11. **Convergence rate of BA-ABC on different dataset.**

competitive performance. Their F-1 scores are similar to those of SVM over various datasets. While SVM often leads in performance, the effectiveness of KNN-Cubic and the Medium Neural Network is particularly evident in datasets

like WHU-RS19 and UC Merced. Overall finding highlights the importance of selecting classifiers tailored to dataset characteristics.

Moreover, the harmonious distribution of Recall, Specificity, and Precision scores across the models points to their overall effectiveness in making a delicate balance between identifying true positives and maintaining accuracy. This illustrates their general competence in differentiating between the classes with a fine level of precision. The detection speed was also compared, e.g. for WHU-RS19, the BA-ABC took 16.32 sec to generate the final result after 60 iterations, while WOA, GWO, BA, ABC, and ACO took 16.09, 16.57, 26.62, 25.26, and 582.90 seconds. This shows that the BA-ABC not only has outstanding accuracy but has excellent detection speed as well.

1) STATISTICAL ANALYSIS OF BA-ABC RESULTS

In this phase of experiments, statistical analysis is carried out on the BA-ABC algorithm using the ANOVA function on five benchmark datasets. Again in these experiments, the same three classifiers (SVM, KNN, and Medium NN) were employed separately for five datasets. The Shapiro-Wilk test p-values of 0.9998, 0.9998, and 1.0 for AID; 1.0, 0.9998, and 0.9998 for RSSCN7; 1.0, 1.0, and, 1.0 for SIRI-WHU; 0.9998, 0.9998, and 1.0 for UC Merced, and 0.9998, 0.9998, and 0.9998 for WHU-RS19 datasets, respectively. All of these values are greater than α, indicating that the data has a normal distribution and homogeneous variances. Therefore, the null hypothesis (H_0) is rejected based on these values.

The ANOVA test results, depicted in tables 24, 25, 26, 27, and 28 suggest that no statistically prominent differences exist among the groups means being compared. For the AID dataset, with an F-statistic of 0.704 and a p-value of 0.531, the data does not provide sufficient evidence to conclude that the group means differ from each other more than what would be expected by chance alone. This indicates that any observed variations in the group means are probably more likely to be the result of random variation than of a systematic distinction between the groups. Similar arguments hold for ANOVA results on other datasets.

Since the p-values are greater than the specified α, we cannot reject the null hypothesis in this scenario. This proves the input data, i.e. the accuracy values of BA-ABC algorithm, have means in a confined range. This proves that the BA-ABC accuracies are consistent across datasets and classifiers. Subsequently, we conduct the ANOVA with Bonferroni for pairwise comparisons.

The Bonferroni post-hoc test is a method used to determine significant differences between multiple groups in an analysis of variance (ANOVA). This test involves conducting pairwise comparisons between all possible pairs of group means to identify specific pairs that differ significantly from each other. Each pairwise comparison is evaluated using a statistical test, such as a t-test, and the resulting p-values are compared to the adjusted significance level.

If a pairwise comparison's p-value is lower than the adjusted significance level, it indicates a significant difference between the corresponding groups. In multiple comparisons, the Bonferroni adjustment aids in reducing the total Type I error rate by making it more stringent. The results of the Bonferroni post-hoc test provide insights into which specific group pairs exhibit significant differences, aiding in a deeper understanding of the data and facilitating more elaborated interpretations of group comparisons in ANOVA analyses.

These tables confirm that the ANOVA and the Bonferroni posthoc test results have $p > \alpha$. Based on the results of pairwise comparisons delineated in Tables 29, 30, 31, 32, and 33, we can assert with 95% confidence that the mean accuracy falls within a narrow range. This proves that our proposed algorithm, BA-ABC consistently performs well across all datasets, regardless of the chosen classifiers. Hence, our proposed approach, employing BA-ABC with the three classifiers, demonstrates significantly superior performance compared to conventional methods. Further, Fig. 10 illustrates the effectiveness of these three classifiers across the five datasets.

2) CONVERGENCE RATE

A convergence rate curve is a graphical representation illustrating the convergence behavior of an optimization algorithm over iterations. It typically plots a performance metric, such as objective function value, against the number of iterations. The curve provides valuable insights into the algorithm's efficiency and effectiveness in reaching the optimal solution.

Rapid convergence is indicated by a fast decrease in the values, which implies that the algorithm is getting close to the ideal solution quite soon. Figure 11 shows the plot of the convergence rate of the BA-ABC.

The curve displays the so-far-best value at each iteration. From the figure, it is evident that BA-ABC exhibits a favorable convergence rate, swiftly identifying a promising region in under 30 iterations.

VI. CONCLUSION

XcelNet17 – a novel deep learning architecture for remote sensing image classification is proposed. It is a rather simple network comprising fourteen convolutional and three fully connected layers, yet demonstrates remarkable performance on five benchmark datasets including AID, RSSCN7, SIRI-WHU, UC Merced, and WHU RS-19. XcelNet17 manages to achieve an Overall Accuracy of 94.6% ∼ 99.9%; substantially surpassing the performance of some of the well-established models such as AlexNet, VGG16, VGG19, ResNet50, and DarkNet19 for the same datasets and three benchmark classifiers. For instance, the proposed architecture yields around 5% improved accuracy on the UC Merced dataset.

Additionally, a meta-heuristic named BA-ABC, which combines the strengths of BA and ABC algorithms, is proposed to improve the features selection process. Assisted by

chaotic mapping for a diverse initial population, BA-ABC demonstrates a significant improvement in the classification accuracy when compared with WOA, GWO, BA, ABC, and ACO algorithms. For example, an 8% superior performance on the WHU-RS19 dataset has been recorded. Various other metrics such as Precision, F1-measure, Recall, and Specificity are also used for the evaluation purposes. To support our research findings, an in-depth statistical analysis has also been presented.

As a follow-up work, we are utilizing vision transformers (ViT) along with CNNs for image classification. While, the CNNs have proven excellent at extraction of local features, they exhibit constrained ability when it comes to global features' extraction, which will be aided by incorporating the ViTs. In addition, we will enhance the framework with another novel nature-inspired feature selection algorithm to further boost the classification accuracy.

REFERENCES

[1] M. A. Moharram and D. M. Sundaram, "Land use and land cover classification with hyperspectral data: A comprehensive review of methods, challenges and future directions," *Neurocomputing*, vol. 536, pp. 90–113, Jun. 2023.

[2] W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review," *Neural Comput.*, vol. 29, no. 9, pp. 2352–2449, Sep. 2017.

[3] A. F. K. Guy, T. Akram, B. Laurent, S. R. Naqvi, M. M. Alex, and N. Muhammad, "A deep heterogeneous feature fusion approach for automatic land-use classification," *Inf. Sci.*, vol. 467, pp. 199–218, Oct. 2018.

[4] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 2, pp. 129–137, Mar. 1982.

[5] L. Ladha and T. Deepa, "Feature selection methods and algorithms," *Int. J. Comput. Sci. Eng.*, vol. 3, no. 5, pp. 1787–1797, 2011.

[6] R. Leardi, "Genetic algorithms in feature selection," in *Genetic Algorithms in Molecular Modeling*. Amsterdam, The Netherlands: Elsevier, 1996, pp. 67–86.

[7] K. Z. Mao, "Orthogonal forward selection and backward elimination algorithms for feature subset selection," *IEEE Trans. Syst., Man Cybern. B, Cybern.*, vol. 34, no. 1, pp. 629–634, Feb. 2004.

[8] J. C. Ang, A. Mirzal, H. Haron, and H. N. A. Hamed, "Supervised, unsupervised, and semi-supervised feature selection: A review on gene selection," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 13, no. 5, pp. 971–989, Sep. 2016.

[9] M. Dorigo, V. Maniezzo, and A. Colorni, "Ant system: Optimization by a colony of cooperating agents," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 26, no. 1, pp. 29–41, Feb. 1996.

[10] D. Karaboga and B. Basturk, "A powerful and efficient algorithm for numerical function optimization: Artificial bee colony (ABC) algorithm," *J. Global Optim.*, vol. 39, no. 3, pp. 459–471, Oct. 2007.

[11] X.-S. Yang, "A new metaheuristic bat-inspired algorithm," in *Nature Inspired Cooperative Strategies for Optimization (NICSO)*. Berlin, Germany: Springer, 2010, pp. 65–74.

[12] X.-S. Yang, "Nature-inspired metaheuristic algorithm," in *Int. Symp. Stochastic Algorithms*. Berlin, Germany: Springer, 2009, pp. 169–178.

[13] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Adv. Eng. Softw.*, vol. 69, pp. 46–61, Mar. 2014.

[14] R. E. J. Kennedy, "Particle swarm optimization," in *Proc. Int. Conf. Neural Netw.*, vol. 4, 1995, pp. 1942–1948.

[15] S. Mirjalili and A. Lewis, "The whale optimization algorithm," *Adv. Eng. Softw.*, vol. 95, pp. 51–67, May 2016.

[16] G. Wang, B. Fan, S. Xiang, and C. Pan, "Aggregating rich hierarchical features for scene classification in remote sensing imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 9, pp. 4104–4115, Sep. 2017.

[17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1–9.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[19] G.-S. Xia, J. Hu, F. Hu, B. Shi, X. Bai, Y. Zhong, L. Zhang, and X. Lu, "AID: A benchmark data set for performance evaluation of aerial scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3965–3981, Jul. 2017.

[20] Q. Zou, L. Ni, T. Zhang, and Q. Wang, "Deep learning based feature selection for remote sensing scene classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 11, pp. 2321–2325, Nov. 2015.

[21] B. Zhao, Y. Zhong, G.-S. Xia, and L. Zhang, "Dirichlet-derived multiple topic scene classification model for high spatial resolution remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 2108–2123, Apr. 2016.

[22] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proc. 18th SIGSPATIAL Int. Conf. Adv. Geographics Inf. Syst.*, Nov. 2010, pp. 270–279.

[23] G.-S. Xia, W. Yang, J. Delon, Y. Gousseau, H. Sun, and H. Maître, "Structural high-resolution satellite image indexing," in *Proc. ISPRS TC VII Symp.*, Vienna, Austria, 2010, pp. 298–303.

[24] D. Dai and W. Yang, "Satellite image classification via two-layer sparse coding with biased image representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 1, pp. 173–176, Jan. 2011.

[25] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[26] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 7263–7271.

[27] R. Minetto, M. Pamplona Segundo, and S. Sarkar, "Hydra: An ensemble of convolutional neural networks for geospatial land classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6530–6541, Sep. 2019.

[28] 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.

[29] Q. Wang, S. Liu, J. Chanussot, and X. Li, "Scene classification with recurrent attention of VHR remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1155–1167, Feb. 2019.

[30] Y. Liu, C. Y. Suen, Y. Liu, and L. Ding, "Scene classification using hierarchical Wasserstein CNN," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 5, pp. 2494–2509, May 2019.

[31] Y. Liu, Y. Zhong, and Q. Qin, "Scene classification based on multiscale convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 12, pp. 7109–7121, Dec. 2018.

[32] J. Fang, Y. Yuan, X. Lu, and Y. Feng, "Robust space–frequency joint representation for remote sensing image scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 7492–7502, Oct. 2019.

[33] J. Wang, W. Li, Y. Wang, R. Tao, and Q. Du, "Representation-enhanced status replay network for multisource remote-sensing image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jun. 28, 2023.

[34] S.-L. Yang, W. Weng, G. Rong, and Y.-P. Feng, "Multiple kernel learning based feature selection for process monitoring," in *Proc. IEEE/ACIS 16th Int. Conf. Comput. Inf. Sci. (ICIS)*, May 2017, pp. 809–814.

[35] E. Boyaci and M. Sert, "Video classification based on ConvNet collaboration and feature selection," in *Proc. 25th Signal Process. Commun. Appl. Conf. (SIU)*, May 2017, pp. 1–4.

[36] D. Koller and M. Sahami, "Toward optimal feature selection," in *Proc. 13th Int. Conf. Mach. Learn. (ICML)*, vol. 96, 1996, pp. 1–14.

[37] S. Kashef and H. Nezamabadi-Pour, "An advanced ACO algorithm for feature subset selection," *Neurocomputing*, vol. 147, pp. 271–279, Jan. 2015.

[38] M. H. Nadimi-Shahraki, S. Taghian, and S. Mirjalili, "An improved grey wolf optimizer for solving engineering problems," *Expert Syst. Appl.*, vol. 166, Mar. 2021, Art. no. 113917.

[39] Y. Zhang, D. Gong, Y. Hu, and W. Zhang, "Feature selection algorithm based on bare bones particle swarm optimization," *Neurocomputing*, vol. 148, pp. 150–157, Jan. 2015.

[40] R. Y. M. Nakamura, L. A. M. Pereira, K. A. Costa, D. Rodrigues, J. P. Papa, and X.-S. Yang, "BBA: A binary bat algorithm for feature selection," in *Proc. 25th SIBGRAPI Conf. Graph., Patterns Images*, Aug. 2012, pp. 291–297.
[41] A. M. Taha, A. Mustapha, and S.-D. Chen, "Naive Bayes-guided bat algorithm for feature selection," *Scientific World J.*, vol. 2013, Jan. 2013, Art. no. 325973.
[42] S. Jeyasingh and M. Veluchamy, "Modified bat algorithm for feature selection with the Wisconsin diagnosis breast cancer (WDBC) dataset," *Asian Pacific J. Cancer Prevention*, vol. 18, no. 5, pp. 1257–1264, May 2017.
[43] S. Kumari Nayak, "Nature inspired algorithms in dynamic task scheduling: A review," *World J. Adv. Res. Rev.*, vol. 20, no. 3, pp. 829–833, Dec. 2023.
[44] P. K. Mandal, "A review of classical methods and nature-inspired algorithms (NIAs) for optimization problems," *Results Control Optim.*, vol. 13, Dec. 2023, Art. no. 100315.
[45] M. G. Bindu and M. K. Sabu, "A hybrid feature selection approach using artificial bee colony and genetic algorithm," in *2020 Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA)*. Cochin, India: IEEE Press, 2020, pp. 211–216.
[46] Y. Saeys, I. Inza, and P. Larrañaga, "A review of feature selection techniques in bioinformatics," *Bioinformatics*, vol. 23, no. 19, pp. 2507–2517, Oct. 2007.
[47] M. Munirah, R. Mohamed, N. M. Nawi, N. Wahid, and M. Shukran, "A comparative analysis on feature selection techniques for classification problems," *ARPN J. Eng. Appl. Sci.*, vol. 11, pp. 13176–13187, Jan. 2016.
[48] C. Shang, M. Li, S. Feng, Q. Jiang, and J. Fan, "Feature selection via maximizing global information gain for text classification," *Knowl.-Based Syst.*, vol. 54, pp. 298–309, Dec. 2013.
[49] J.-H. Lee, J. Rahimipour Anaraki, C. W. Ahn, and J. An, "Efficient classification system based on Fuzzy–Rough feature selection and multitree genetic programming for intension pattern recognition using brain signal," *Expert Syst. Appl.*, vol. 42, no. 3, pp. 1644–1651, Feb. 2015.
[50] G.-G. Wang and Y. Tan, "Improving metaheuristic algorithms with information feedback models," *IEEE Trans. Cybern.*, vol. 49, no. 2, pp. 542–555, Feb. 2019.
[51] S. Gupta and K. Deep, "A novel random walk grey wolf optimizer," *Swarm Evol. Comput.*, vol. 44, pp. 101–112, Feb. 2019.
[52] A. E. Hegazy, M. A. Makhlouf, and G. S. El-Tawel, "Improved salp swarm algorithm for feature selection," *J. King Saud Univ., Comput. Inf. Sci.*, vol. 32, no. 3, pp. 335–344, Mar. 2020.
[53] G. I. Sayed, A. Tharwat, and A. E. Hassanien, "Chaotic dragonfly algorithm: An improved metaheuristic algorithm for feature selection," *Appl. Intell.*, vol. 49, no. 1, pp. 188–205, Jan. 2019.
[54] G. C. Dandy, A. R. Simpson, and L. J. Murphy, "An improved genetic algorithm for pipe network optimization," *Water Resour. Res.*, vol. 32, no. 2, pp. 449–458, Feb. 1996.
[55] D. Rodrigues, L. A. M. Pereira, R. Y. M. Nakamura, K. A. P. Costa, X.-S. Yang, A. N. Souza, and J. P. Papa, "A wrapper approach for feature selection based on bat algorithm and optimum-path forest," *Expert Syst. Appl.*, vol. 41, no. 5, pp. 2250–2258, Apr. 2014.
[56] P. Manchala, M. Bisi, and S. Agrawal, "BAFS: Binary artificial bee colony based feature selection approach to estimate software development effort," *Int. J. Inf. Technol.*, vol. 15, no. 6, pp. 2975–2986, Aug. 2023.
[57] Y. Jiang, T. Hu, C. Huang, and X. Wu, "An improved particle swarm optimization algorithm," *Appl. Math. Comput.*, vol. 193, no. 1, pp. 231–239, 2007.
[58] M. Schiezaro and H. Pedrini, "Data feature selection based on artificial bee colony algorithm," *EURASIP J. Image Video Process.*, vol. 2013, no. 1, pp. 1–8, Dec. 2013.
[59] A. Tamiru and F. Hashim, "Application of bat algorithm and fuzzy systems to model exergy changes in a gas turbine," in *Artificial Intelligence, Evolutionary Computing and Metaheuristics: In the Footsteps of Alan Turing*. Berlin, Germany: Springer, 2013, pp. 685–719.
[60] T. C. Bora, L. d. S. Coelho, and L. Lebensztajn, "Bat-inspired optimization approach for the brushless DC wheel motor problem," *IEEE Trans. Magn.*, vol. 48, no. 2, pp. 947–950, Feb. 2012.
[61] A. H. Gandomi, X.-S. Yang, A. H. Alavi, and S. Talatahari, "Bat algorithm for constrained optimization tasks," *Neural Comput. Appl.*, vol. 22, no. 6, pp. 1239–1255, May 2013.
[62] D. Karaboga, "An idea based on honey bee swarm for numerical optimization," Dept. Fac. Comput. Eng., Erciyes Univ., Kayseri, Turkey, Tech. Rep. tr06, 2005.
[63] L. Stahle and S. Wold, "Analysis of variance (ANOVA)," *Chemometrics Intell. Lab. Syst.*, vol. 6, no. 4, pp. 259–272, Nov. 1989.
[64] T.-T. Wong, "Parametric methods for comparing the performance of two classification algorithms evaluated by k-fold cross validation on multiple data sets," *Pattern Recognit.*, vol. 65, pp. 97–107, May 2017.
[65] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, no. 3, p. 591, Dec. 1965.

BILAL AHMED received the B.Sc. and M.Sc. degrees in electrical engineering from the University of Engineering and Technology, Taxila, Pakistan, in 2005 and 2012, respectively. He is currently a Ph.D. Researcher with the Department of Electrical and Computer Engineering, COMSATS University Islamabad, Wah Campus. He has been actively involved in various research areas. His research interests include embedded systems, computer vision, artificial intelligence, and machine learning applications. He has more than 18 years of industrial experience in these areas.

TALLHA AKRAM received the B.Sc. degree in computer engineering from COMSATS University Islamabad, Abbottabad Campus, Pakistan, in 2006, the M.Sc. degree in embedded systems and control engineering from Leicester University, U.K., in 2008, and the Ph.D. degree in computer vision and pattern recognition from Chongqing University, China, in 2014. He is currently an Associate Professor with the Electrical and Computer Engineering Department, COMSATS University Islamabad, Wah Campus, Pakistan. He is the author of a number of peer-reviewed journals and conferences. His research interests include computer vision, pattern recognition, machine learning, artificial intelligence, and applied optimization.

SYED RAMEEZ NAQVI received the M.Sc. degree in electronic engineering from The University of Sheffield, U.K., in 2007, and the Ph.D. degree in computer engineering from Vienna University of Technology, Austria, in 2013. He is currently an Associate Professor with the Department of Electrical and Computer Engineering, COMSATS University Islamabad, Wah Campus. He has published more than 40 research articles in international journals and symposiums of high reputation and carried out several funded research projects with various research and development organizations. His research interests include asynchronous logic, computer architecture, hardware optimization, and artificial intelligence.

ANAS ALSUHAIBANI received the M.Sc. degree in information systems and the Ph.D. degree in information studies from the Information School, The University of Sheffield, U.K., in 2017 and 2022, respectively. He is currently an Assistant Professor with the Information Systems Department, Prince Sattam Bin Abdulaziz University, Saudi Arabia. His research interests include artificial intelligence, machine learning, computer vision, and natural language processing.

YOUSSEF N. ALTHERWY received the Ph.D. degree in computer science from Imperial College London, U.K., in 2022. He is currently an Assistant Professor with the Department of Information Systems, College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University. His research interests include wireless sensor networks (WSNs), machine learning, signal processing, and data science.

USMAN MASUD received the graduate degree in electrical engineering from the University of Engineering and Technology, Taxila, followed by two postgraduate degrees in electrical engineering from National University of Sciences and Technology, Rawalpindi, and University of Kassel, Kassel, where he finalized his Ph.D. degree (by research work) at the latter university. He has supervised and completed multiple funded projects at the national and international levels and finds deep interest in laser-based biomedical applications. His research focus includes optoelectronic systems, image processing, biomedical sensors, machine learning, spectroscopic applications, AI, computer vision, neural networks, and wireless networks. He has been an active member of numerous professional networks like Verband der Elektrotechnik, Elektronik und Informationstechnik e.V. (VDE) for several years.