JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Quasi Oppositional based Artificial Fish Swarm Optimization Algorithm with Mobile Net for Remote Sensing Image Scene Classification

N.Bharanikumar[1*], Dr. P. Dhanalakshmi[2]

[1*] Department of Computer and Information Science, Annamalai University, Annamalai Nagar 608002 E-mail: cdmbharani81@gmail.com

[2] Department of Computer Science & Engg., Annamalai University, Annamalai Nagar, 608002 E-Mail: abidhana01@gmail.com

Abstract

Presently, the domain of unmanned aerial vehicles (UAVs) has gained significant interest in several application areas owing to its characteristics of inexpensive, versatile, and autonomous. Robust remote sensing scene classification process find useful in several UAV based surveillance systems like forest fire detection, which facilitating object detection, tracking process and drastically improves the result of visual surveillance. The recently developed deep learning (DL) models can be employed for automated remote sensing scene classification. This paper presents an effective Quasi Oppositional based Artificial Fish Swarm Optimization Algorithm with MobileNet(QAFSOMN) model for aerial image classification. The presented QAFSOMN model uses MobileNet based feature extraction technique, which is further optimized by the QAFSO algorithm to optimally select the hyperparameters. In addition, two classification models such as gradient boosting tree (GBT) and random forest (RF) are used to determine the different classes of remote sensing images. The utilization of the QAFSO algorithm paves a way of achieving higher detection performance compared to the existing methods.The detailed experimental validation process on remote sensing imagery dataset pointed out that the QAFSOMN-GBT and QAFSOMN-RF models have obtained a maximum accuracy of 0.9871 and 0.9784.

Keywords: Deep learning, Machine learning, Hyperparameter tuning, Remote sensing classification, Feature extraction

1. Introduction

In recent times, Unmanned Aerial Vehicles (UAVs) have gained maximum concentration among the developers which is applied in search-and-rescue, monitoring, reconnaissance, and analysis of structure [1]. Land-cover classification is one of the important units of UAV domains which is highly complicated for developing independent systems. The object prediction operation is extremely composite, and requirement cost

is highly challenging. Based on the movement of UAV, images are blurred, with noisy details as on-board cameras are prominently generated with low-resolution images. In enormous domains of UAVs, prediction operation is complicated because of the requirement in real-time performance. In addition, numerous studies on UAV deal with monitoring certain objects, like vehicles, landmarks, landing areas, and users including trespassers. Hence, few works have been considered for object prediction as multiple target object forecasting is significant for various UAV sectors.

The space among the technical capabilities due to the existence of 2 significant limitations: (i) it is difficult to develop and record diverse approaches of target objects and (ii) massive processing power is essential for real-time object prediction even under a single object. Aerial imagery classification of scenes has retrieved as sub-regions by covering enormous ground objects and classes of land-cover as diverse semantic categories. Followed by, for practical domains of remote sensings, such as urban planning, computer cartography, and resource management, aerial image categorization is extremely important [2]. In general, a similar object classes of land cover are distributed among diverse scenarios. For example, residential as well as commercial are 2 main classes of scenes with trees, constructions, and roads; but, the variations among spatial distribution and density of above-mentioned categories. Hence, in aerial scenarios, classification relied on structural as well as spatial pattern complexity is one of the major issues.

Basically, this model is to develop a holistic scene implication for scene classification. In remote sensing research, a significant technology used for scene classification is bag-of-visual-words (BoVW). The main purpose of BoVW is to examine the text which develops a document by sequential words. For recognizing images by count of VW which is produced by local feature quantizing, bag of words (BOW) framework has been employed along with a clustering scheme. BoVW technology is an extended version of BoW for image analysis, in which an image is described as VW words from visible dictionary by histogram of VW.

Followed by, Deep Learning (DL) models [3] were applicable to resolve the existing issues such as object prediction and analysis, natural language computation, and speech analysis, and in massive practical domains. Besides, this method is highly effectual when compared with standard principles, and it has gained better attention in industrial and academics applications. The DL scheme manages to gain general hierarchical properties for learning interms of diverse abstraction levels. Deep convolutional neural networks (CNNs) [4] are typical DL method. In recent times, a prominent and effective one in prediction and examination operations which provides better outcomes. In case of image classification, CNN is well-known in past decades as it enables to accomplishment supreme classification accuracy. In case of industrial application, it is impossible to apply a proficient deep CNN as the difficulty of manual fine-tuning the hyper-parameters as well as trade-offs between processing cost and classification accuracy. This work is presented for reducing the processing cost. In UAV aerial scene categorization, difficulty of remarkable CNNs was limited. A special type of CNN architecture is decided for reducing the search space and minimum search space is equated under the application of particular domain knowledge.

This paper presents an effective Quasi Oppositional based Artificial Fish Swarm Optimization Algorithm based MobileNet(QAFSOMN) model for aerial image classification. The presented QAFSOMN model utilizes MobileNet based feature extractor to derive an optimal feature vector. Followed by, the hyperparameters of the

MobileNet are tuned by the QAFSO algorithm to optimally select the hyperparameters. In addition, two classification models such as gradient boosting tree (GBT) and random forest (RF) is used to determine the dissimilar classes of remote sensing images. A set of simulations were carried out to ensure the superior classification performance on the remote sensing images.

2. Related works

Massive works were presented as extended models of BOW relied on specific characteristics of aerial scenes. [5] established a spatial co-occurrence kernel (SCK), which defines spatial distribution of VW. [6] proposed a transformation and rotation-invariant pyramid-of-spatial-relations (PSR) scheme for defining the relative and absolute spatial relationships of local properties. [7] modeled a concentric-circle-structured multi-scale BOW (CCM-BOW) approach which has been employed for accomplishing rotation invariance. Then, [8] deployed a rotation-invariant scheme using collections of part detectors (COPD) that gains visible parts of images. Even though it has obtained supreme performance, where it is evolved from classical BOW approaches, and it is impossible to enhance the outcomes of limited capability of low-and mid-level features.

Unsupervised feature learning (UFL) is a well-known framework and Machine Learning (ML) researchers. The UFL technique learns the features from huge unknown instances independently. It is able to find the related data acquired by corresponding data. [9] applied a sparse coding system for learning sparse local features from image scenarios and for collecting local parameters for generating the image representation. [10] exploits a previous NN named as a sparse autoencoder (SAE), which has trained a collection of decided image patches that has been sampled under the application of saliency degree for extracting local properties. [11] improved the existing UFL pipeline with the help of known features. Even though UFL framework are hand-engineered features, where enhancement of shallow learning structures. The popularity of CNN is applicable in various domains.

[12] initialized a scheme for CNN training under the application of backpropagation (BP) mechanism and reached better performance in character analysis. In recent times, CNN was applied prominently in computer vision tasks. Simultaneously, it is tedious to train deep CNN with massive number of parameters. [13] implied better outcomes by investigating the texture by using pooling CNN variables accomplished from convolutional layers and Fisher coding process. The application of CNNs in scene classification is still a challenging issue. In [14], a pre-defined CNN was employed and tuned completely on a scene dataset which has showcased supreme outcomes while pre-trained CNN is converted as scene datasets which lack in training modules. In [15], a generalized potential of CNN features acquired from fully connected (FC) layers are sampled while categorizing remote sensing images, and yielded optimal results when compared with open access scene dataset.

3. The Proposed Model

The workflow involved in the QAFSOMN model with GBT and RF classifiers are shown in Fig. 1. Primarily, the input aerial image is pre-processed to boost up the quality of the image. Afterward, QAFSOMN model is applied as a feature extraction technique to derive an actual set of feature vectors. In order to raise the convergence rate of the QAFSO algorithm, quasi oppositional based learning (QOBL) concept is applied. At last, the feature vectors are fed to the GBT and RF models to perform classification task.

3.1. Preprocessing

The remote sensing images collected by the UAVs might not be clear in some scenarios. When the object color is not static, image segmentation approaches enable the segmentation of frames into target areas and irrelevant objects. It enables the frame region to be processed and also discard the frames with no probable objects. Using a sliding window, in the HSV color space, a frame undergo scanning in the preprocessing stage, and all windows are verified by the saturation window component thresholding $th_{sat}(V)$. The thresholding is carried out using Eq. (1)

$$th_{sat}(V) = 1.0 - \frac{0.8V}{255} \quad (1)$$

where V is the intensity component rate. Once the saturation component rate exceeds to $th_{sat}(V)$, the pixels are computed based on the object.

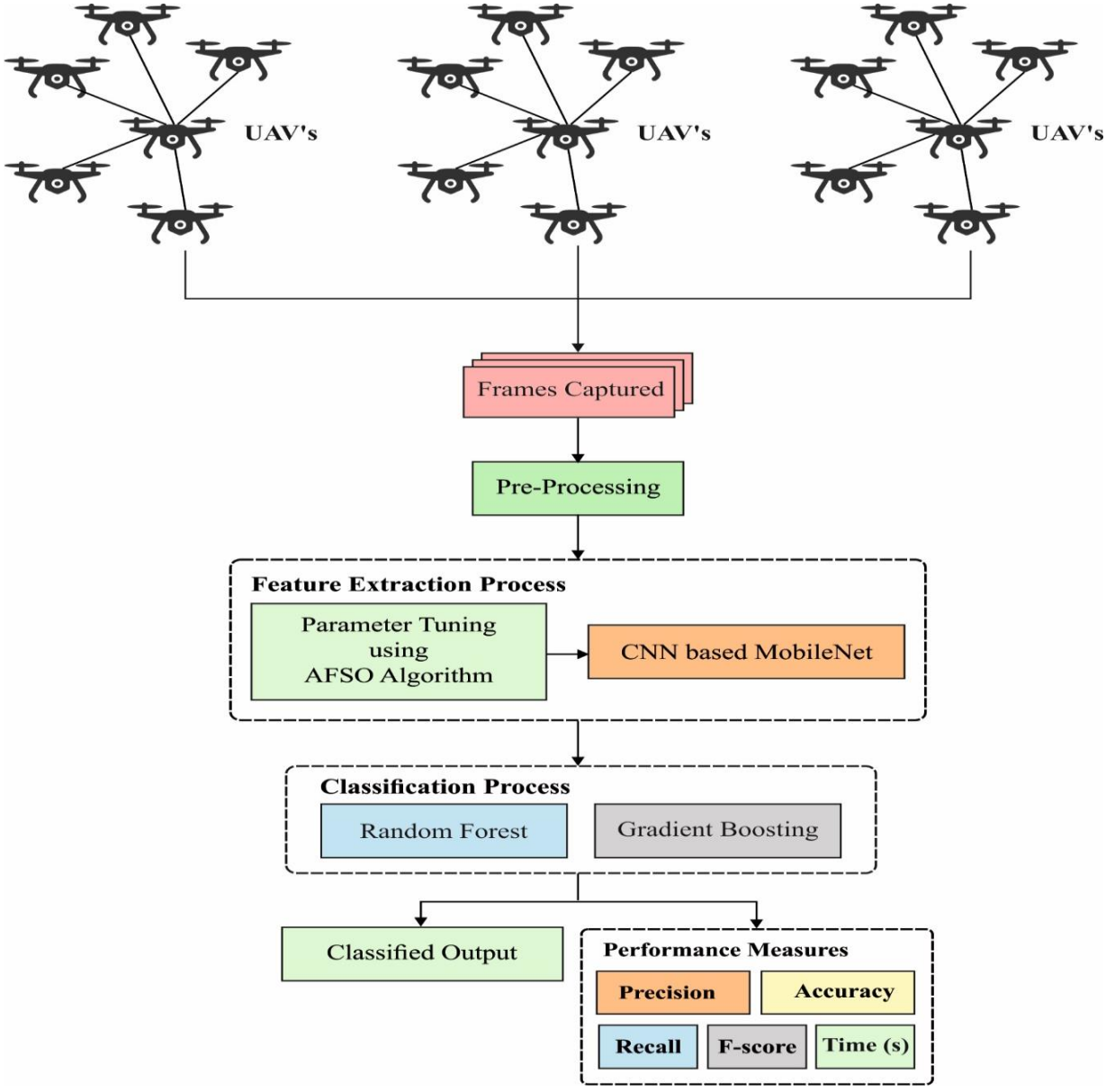

Fig. 1. Architecture of Proposed Methods

3.2. Feature Extraction

The MobileNetsystem was deployed for enhancing the practical function of DL models with minimum hardware applications. It is capable of reducing the attribute count with no limited accuracy.Traditionally, the

MobileNethas offered partial number of parameters under VGG-16for gaining similar classification accuracy in ImageNet-1000 classification tasks.

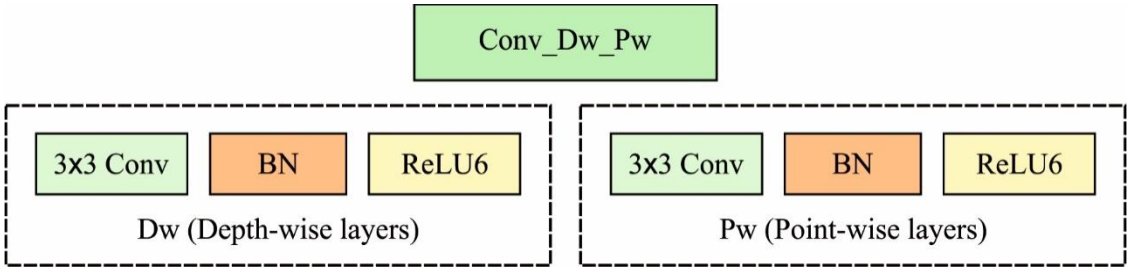

Fig. 2. Structure of MobileNet

Fig. 2 implied the fundamental convolution structure of MobileNet. $Conv_Dw_Pw$is one of the deep and differentiable convolution infrastructure. It is comprised of depth-wise layers (Dw) and point-wise layers (Pw). Here, Dwis a deep convolutional layer which has applied 3×3 kernels, whereas Pw is convolutional layers which help 1×1 kernels. A convolution outcome are facilitated using BN approach and activation function named as rectified linear unit (ReLU).

Here, activation function ReLU is interchanged by $ReLU6$, and normalization is performed under the application of Bayesian Network (BN) framework which supports automated extension of data distribution. Hence, $ReLU6$is demonstrated as:

$$y = \min(\max(z, 0), 6) \tag{2}$$

where z means the measure of a pixel in feature map.The deep as well as separable convolutional architecture activates MobileNet to enhance training and limits the volume of estimation. The remarkable convolution is defined as:

$$G_N = \sum_M K_{M,N} * F_M \tag{3}$$

where$K_{M,N}$denotes a filter; and M and N are count of input channels as well as output channels. In standard convolution, input image, with feature image, $F_M$implies input images with feature maps, which applies style of zero paddings. Fig. 3 shows the layered architecture of MobileNet.

Type / Stride	Filter Shape	Input Size
Conv / s2	3 x 3 x 3 x 32	224 x 224 x 3
Conv dw / s1	3 x 3 x 32 dw	112 x 112 x 32
Conv / s1	1 x 1 x 32 x 64	112 x 112 x 32
Conv dw / s2	3 x 3 x 64 dw	112 x 112 x 64
Conv / s1	1 x 1 x 64 x 128	56 x 56 x 64
Conv dw / s1	3 x 3 x 128 dw	56 x 56 x 128
Conv / s1	1 x 1 x 128 x 128	56 x 56 x 128
Conv dw / s2	3 x 3 x 128 dw	56 x 56 x 128
Conv / s1	1 x 1 x 128 x 256	28 x 28 x 128
Conv dw / s2	3 x 3 x 256 dw	28 x 28 x 256
Conv / s1	1 x 1 x 256 x 256	28 x 28 x 256
Conv dw / s1	3 x 3 x 256 dw	28 x 28 x 256
Conv / s1	1 x 1 x 256 x 512	14 x 14 x 256
5 x Conv dw / s1	3 x 3 x 512 dw	14 x 14 x 512
5 x Conv / s1	1 x 1 x 512 x 512	14 x 14 x 512
Conv dw / s2	3 x 3 x 512 dw	14 x 14 x 512
Conv / s1	1 x 1 x 512 x 1024	7 x 7 x 512
Conv dw / s2	3 x 3 x 1024 dw	7 x 7 x 1024
Conv / s1	1 x 1 x 1024 x 1024	7 x 7 x 1024
Avg Pool / s1	Pool 7 x 7	7 x 7 x 1024
FC / s1	1024 x 7	1 x 1 x 1024
GNB, SVM / s1	Classifier	1 x 1 x 7

Fig. 3. Layered Architecture of MobileNet

If the size and channels of input images are $D_F * D_F$ and M, then it is essential to have N filters with M channels as well as size of $D_K * D_K$in prior to generating$N$ feature images of size $D_K * D_K$. Therefore, processing cost is $D_K * D_K * M * N * D_F * D_F$. Unlike, Dwis formulated as:

$$\hat{G}_M = \sum \hat{K}_{1,M} * F_M \tag{4}$$

where$\hat{K}_{1,M}$refers a filter. $F_M$has similar meaning as Formula (4). If the step size is 1, then zero assures that size of a characteristic graph is invariable after using deep as well as separable convolutional architecture. If the step size is 2, zero fillings assure that size of feature graph accomplished after using deep and separable convolutional infrastructure is a partial input image/feature graph; which is a dimensional reduction task as well.

The deep separable convolution architecture of MobileNetis capable of gaining identical results where remarkable convolution depends upon similar inputs. The Dw phase requires M filters with a channel and size of $D_K * D_K$. Followed by, Pw phase requires N filters using M channels and size of 1×1. Therefore, processing cost of deep separable convolution architecture is $D_K * D_K * M * D_F * D_F + M * N * D_F * D_F$,

where $N1 + \frac{1}{D_K^2}$ of a standard convolution. Next, data distribution might be altered by a convolution layer at the time of network training. When data is present on the edge of activation function, then a gradient has vanished where it is not upgraded. Likewise, the standard normal distribution, BN model extends a data by allocating 2 learning attributes and eliminates gradient vanishing and extension of complicated parameters.

3.3. QAFSO algorithm-based parameter tuning

The hyperparameters involved in the QAFSO algorithm is determined using the QAFSO algorithm. The incorporation of QOBL concept helps to increase the convergence rate of AFSO algorithm. AFSO is a well-known Swarm Intelligent (SI) optimization model which depends upon the behavior of fish-swarming. This model is highly beneficial for identifying global optimal solution and does not acquire gradient details of the objective function. Here, an artificial fish explores the food according to the foraging hierarchy of swarming nature as well as random behavior.In addition, the artificial fish enables mutual data communications till reaching a global optimum.

The fundamental concept of AFSOis defined in the following: an *n*-dimensional space, assume that a fish swarm with N artificial fish. Consider $X = (x_1, x_2, \dots x_n)$means the place of artificial fish, and $Y = f(X)$denotes the fitness at Position X. Assume $d_{ij} = \|X_i - X_j\|$is a distance among the position $X_i$and X_j, and *Visual* and *Step*implies the perception range and moving step of artificial fish, correspondingly.

(1) Foraging behavior

Consider $X_i$is a recent state of artificial fish, and choose the state Xjarbitrarily from *Visual* range. When $Yj < Yi$, then artificial fish is moved a *Step* in direction of $(X_j - X_i)$. Else, decide a state Xjin random fashion for selecting whether it meets the forward criteria. When the condition is not satisfied, then random behavior is carried out.

The foraging nature applies the given rule:

$$\tilde{X}_i = \begin{cases} X_i + Step \cdot \dfrac{X_j - X_j}{d_{jj}} \cdot rand, & if\ (Y_j < X_i) \\ randombehavior, & otherwise \end{cases} \quad (5)$$

where$\tilde{X}_i$means the upcoming state of anartificial fish, *rand*denotes uniformly produced values from [0,1].

(2) Swarming behavior

In a fish swarm, artificial fish $X_i$has to search intermediate place X_c of N_F artificial fish in recent neighborhood(d_{ij} <Visual). If $(Y_C/N_F > \delta Y_i)$, the artificial fish $X_i$moves forward into X_c. The numerical function of swarming behavior is provided below:

$$\tilde{X}_i = \begin{cases} X_i + Step \cdot \dfrac{X_c - X_i}{d_{ic}} \cdot rand, & if\ (Y_c/N_F < \delta \cdot Y_i) \\ foragyingbehavior, & otherwise \end{cases} \quad (6)$$

where$\delta \in (0,1)$denotes the food concentration.

(3) Following behavior

If X_{lbest} is a local best unit in present neighborhood of X_i. Then, $(Y_{lbest}/N_F > \delta Y_i)$, the artificial fish $X_i$moves in a direction $(X_{lbest} - X_i)$. The arithmetic expression of this behavior is implied as:

$$\tilde{X}_i = \begin{cases} X_i + Step \cdot \dfrac{X_{lbest} - X_i}{d_{i,lbest}} \cdot rcmd, & if\ (Y_{lbest}/N_F < \delta \cdot Y_i) \\ foragingbehavior, & otherwise \end{cases} \quad (7)$$

(4) Random behavior

The artificial fish decides a place randomly from *Visual* range and travels to the respective place. It is named as a default behavior.

(5) Behavior selection

In case of AF, predefined behaviour is performed and compared, correspondingly. Therefore, an optimal nature has been decided for upgrading recent state of AF.

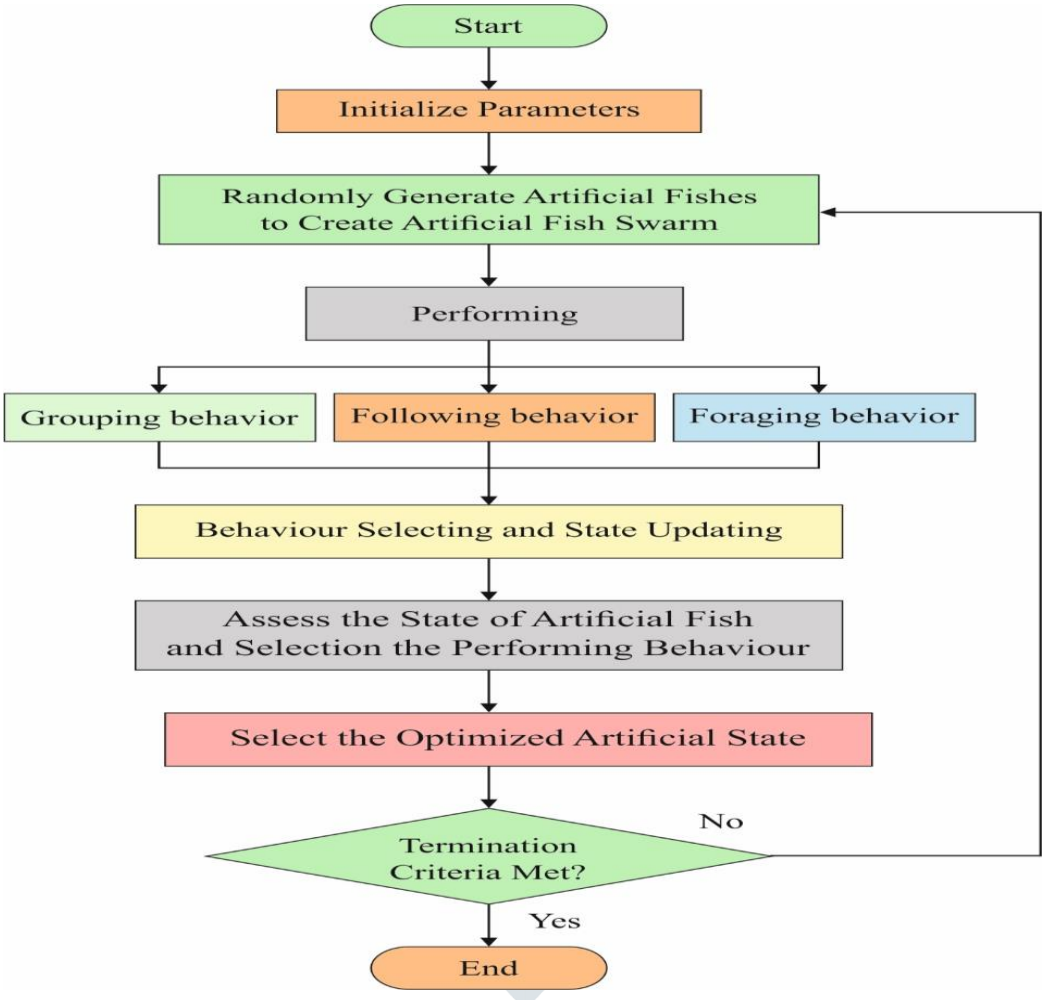

Fig. 4. Flowchart of AFSA model

(6) Bulletin

It is applied for recording best state X_{best} in a fish swarm. Every AFs are compared with corresponding state. When the condition becomes normal, then a bulletin might be upgraded. Therefore, AFSO applies a social nature of fish swarm to resolve the optimization issues, and it is extremely beneficial for fish self-information as well as environmental data for changing the searching direction in order to gain better diversity and convergence. Hence, AF gets a position where the food resource is maximum. Even though AFSO algorithm is supreme in global optimization model for optimization issues, it is still risk in converging sub-optimum such as metaheuristics. It is named as premature convergence of complex optimization issues which results in reduced efficiency.For improving the convergence rate of AFSO algorithm, QOBL concept is introduced for population initialization. Rahnamayan et al.[16] initially developed OBL, including that quasi opposite numbers have a

maximum chance of attaining a solution compared to random numbers. The combination of metaheuristic algorithms with QOBLenhances the accuracy and convergence rate. Fig. 4 depicts the flowchart of AFSA model.

3.4. Classification

The final step of remote sensing image classification involves GBT and RF models, which receive the feature vectors from the QAFSOMN model and produces appropriate class labels as output.

3.3.1. GBT model

Here, weak classifiers are differentiated and weighted every sample. At the initial stage, similar weightsare allocated for a sample and applied initial sample for the purpose of training weak classifiers. Once the learning process is completed, weight of error samples has been improved and reduce the weight of accurate sample. Followed by, second weak classifier has been applied for learning a data sample. Finally, m weak classifiers are accomplished and after the combination of m classifiers, the last classifier has been achieved.GBT is belonging to the collective methods on the basis of various weak classifiers. Thus, the final outcomes are biased by weighting model. Weak classifiers have applied Regression decision tree (DT) in massive iterations. Each weak classifier undergoes training according to the former weak classifier's error where GBThas reached a classification target by decreasing the error in training process.

Actually, GBT is a type of improving ML model. Hence, the inclusion model and forward stage-wise methods are applied. The given formula showcases the mth weak classifier, where Mimplies count of classifiers and $\Theta_m$exhibits a variable of certain classifiers.

$$f_M(x) = \sum_{m=1}^{M} T\,(x, \Theta_m) \tag{8}$$

The forward stagewise method means iterative model from front to back, learning one weak classifier with parameters. The learning process of present weak classifier depends upon the weak classifiers that has been trained initially. Hence, mth step of boosting classification is depicted as given below.

$$f_m(x) = f_{m-1}(x) + T(x, \Theta_m) \tag{9}$$

Loss Function:

$$L(f_m(x), y) = L(f_{m-1}(x) + T(x, \Theta_m), y) \tag{10}$$

For training process, train a classifier $T(x, \Theta_m)$ in all iterations which ensures Loss Function $L(f_m(x), y)$to be limited.

In general, GBT depends upon Negative binomial logarithm likelihood $\log(1 + e^{-2yF}), y \in (-1, +1)$. It is simple for optimizing loss function, however, it is complicated to optimize a common function using Gradient Descent (GD) scheme. Followed by, Freidman devised a technology that has applied negative gradient of loss function to fit the Classification as well as Regression Tree. The particular strategy is showcased in the following:

Consider the N sample sets and initiate weak classifiers:

$$f_0(x) = \arg\min_{c} \sum_{i=1}^{N} L\,(y_i, c) \tag{11}$$

The tth round ith sample loss's Negative GD is depicted as:

$$r_{ti} = \left[\frac{\partial L\left(y_i, f(x_i)\right)}{\partial f(x_i)}\right]_{f(x)=f_{t-1}(x)} \tag{12}$$

It is applied $(x_i, r_{it})i = 1,2, \ldots m$ to fit tth Decision Regression Tree, the parallel leaf node site is $R_{tj}, j = 1,2, \ldots J$. Jimplies a count ofleafnodes. The sample of a node seeks for resultant value c_{tj} to limit the loss function.

$$c_{tj} = \arg\min \sum_{x_i \in R_{tj}} L\left(y_i, f_{t-1}(x_i) + c\right) \tag{13}$$

Reform round and accomplish a robust classification approach.

$$f_t(x) = f_{t-1}(x) + \sum_{j=1}^{J} c_{tj}\, I\left(x \in R_{tj}\right) \tag{14}$$

Finally, last strong classifier is depicted as,

$$f(x) = f_M(x) = f_0(x) + \sum_{t=1}^{M}\sum_{j=1}^{J} c_{tj}\, I\left(x \in R_{tj}\right) \tag{15}$$

In case of classification issue, the Negative binomial logarithm likelihood $L(y,F) = \log\left(1 + e^{-2yF}\right), y \in (-1, +1)$has been applied. Thus, negative gradient error is illustrated as;

$$r_{ti} = \left[\frac{\partial L\left(y_{i'} f(x_i)\right)}{\partial f(x_i)}\right]_{f(x)=f_{t-1}(x)} = \frac{y_i}{1 + \exp\left(y_i f(x_i)\right)} \tag{16}$$

Thus, Best Residual Error value for all nodes are given below:

$$c_{tj} = \arg\min \sum_{x_i \in R_{tj}} log\left(1 + \exp\left(-y_i(f_{t-1}(x_i) + c)\right)\right) \tag{17}$$

Followed by, majorization that applies approximate measures as a substitute.

$$c_{tj} = \frac{\sum_{x_i \in R_{tj}} r_{ti}}{\sum_{x_i \in R_{tj}} |r_{ti}|(1 - |r_{ti}|)} \tag{18}$$

Grid Search is defined as a basic parameter majorization model which classifies the attributes into grids with identical length in special range of coordinate system. Each point in a coordinate system shows a collection of variables and applies the point in greater extent for validation of model performance in all iterations. Therefore, a point which performs optimally is named as optimal parameter. Followed by, a method of Grid Search is to traverse points belongs to all grids.

3.3.2. RF model

RF is defined as a combined learning model which unifies various DT for the purpose of eliminating correlation among feature data. Simultaneously, processing complexity of RF is $O(n)$ ($n$is No. of samples) if massive is present, moreover, it is implemented in parallel as this mechanism is to enhance the performance. RF limits the correlation among DT by random selection of samples as well as features. Initially, the identical

quantity of data has been decided randomly from training sample. Next, few portion of features are chosen randomly for introducing DT. Such types of randomization makesa correlation among DT and limits the failure which exists if the DT is over-fitting, and enhances the model accuracy.

Gini Coefficient

While the DT is generated, the value of data is described as the maximum the "uncertainty" of data is limited, the more data the partition is accomplished. There are 2 typical indicators used for calculating the uncertainty: data entropy as well as *Gini* index.

Consider K random parameters, the definition of *Gini* coefficient is:

$$Gini\ (y) = 1 - \sum_{k=1}^{K} p_k^2 \tag{19}$$

where$p_k$implies various probabilities of considering Kth variable. It is approved that if the Eq. (2) is met, then higher $Gini(y)$ is attained, and when $p_i = 1$and $p_j = 0, (i \neq j)$, then $Gini(y) = 0$. This implies that the maximum irregular y(y is the variable is defined) is, the larger $Gini(y)$as well. Therefore, *Gini* coefficient could be applied for measuring uncertainty

$$p_1 = p_2 = \cdots = p_K = \frac{1}{K} \tag{20}$$

Decision Tree

The method applied in this work is a CART approach which is a classification and regression tree (CART) under the application of *Gini* gain as terminating condition, for CART is sophisticated and applied for resolving classification and regression issues. The steps involved in the classification task is given below.

- Initialize data set D on a node;
- If all the samples in D belong to the category c_k, the node will not continue to generate and mark it as c_k,
- When there is no feature, the class with massive samples in D is considered as class of a node;
- Else, when feature $x^{Q)}$ has $S_j$diverse measures $u_1^{(j)}, \cdots, u_{s_j}^{(j)}$ which meets $u_1^{(j)} < \cdots < u_{s_j}^{(j)}$ in a recent data set, then:

(i) When $x^{(j)}$ is discrete, $u_1^{(j)}, \cdots, u_{s_j}^{(j)}$ are decided as isolation points $a_p$successively, then:

$$A_{jp} = \{x^{(j)} = a_p, x^{(j)} \neq a_p\} \tag{21}$$

(ii) When$x^{(j)}$ is continuous, $\frac{u_1^{(j)}+u_2^{(j)}}{2}, \cdots, \frac{u_{s_j-1}^{(j)}+u_{s_j}^{(j)}}{2}$ are decided as separation points a_p successively, then:

$$A_{jp} = \{x^{(j)} < a_p, x^{(j)} \geq a_p\} \tag{22}$$

A_{jp}means the result of division of $x^{(j)}$.Based on the *Gini* gain, the feature $x^{(j^*)}$ with maximum IG of jth feature. The parallel dichotomy is estimated as division criteria:

$$(j^*, p^*) = \arg\max_{j,p} g_{Gini}(y, A_{jp}) j, p \tag{23}$$

- Whentermination criteria is met, then consider the class with massive count of samples in Das the final category;

- Else, based on feasible values of $x^{Q^*)}$ ($\{a_1, \dots, a_m$ divide D into $\{D_1, \dots, D_m\}$:

$$(x_i, y_i) \in D_j, \left(x_i^{(j*)} = a_j\right), \forall i = 1, \cdots, N \tag{24}$$

- Repeat the model for each D_j.

By looping the above steps, a DT which satisfies a certain goal is produced.The variations among regression as well as classification tree is a node division criteria of nodes and decision of output. The termination condition is a least-squares model. In case of $x^{(j)}$ scan the viable measures, and decides the isolation point a_p, then $x^{(j)}$is classified as 2 portions R_1 and R_2. Identify the value c_1 and $c_2$from the result y, correspondingly, until a small value is attained. However, this $a_p$defines the optimal separation point of $x^{(j)}$

$$\min_{a_p}\left[\min_{c_1}\sum_{x_i \in R_1}(y_i - c_1)^2 + \min_{c_1}\sum_{x_i \in R_2}(y_i - c_1)^2\right] \tag{25}$$

In line with this, best partitioning properties $x^{(j^*)}$is accomplished by traversing j and parallel nodes.

The output value is computed by average values of corresponding radius. Consider R_1 as a sample, the final value is depicted as:

$$c_o = \frac{1}{N_1}\sum_{x_i \in R_1} y_j \tag{26}$$

Hence, N_1means the count of instances in R_1.

Once the DT is produced, input a sample feature values which has to be computed, the parallel result would be accomplished.

4. Experimental Validation

The newly developed method was performed using a standard remote sensing image dataset [17], with 21 classes of land images and 100 images under each class label. The size of an image is 256*256 pixels. The filtrations of images are carried out manually from the USGS National Map Urban Area Imagery collection with numerous urban regions globally. The resolution of pixels from typical images is 1 foot. Various sample images with a class label are shown in Fig. 5.

Fig. 5. Sample Images from UCM Dataset

Table 1 and Fig. 6 showthe comparative results analysis of presented method with previous DL schemes on the given dataset [18]. The final measures have exhibited the poor classification of the Places Net method with minimum accuracy of 0.9144. In addition, the VGG-VD19 scheme has provided maximum accuracy of 0.9315, which goes beyond the function of Places Net. Simultaneously, 6 DL methodologies like VGG-VD16, VGG-F, AlexNet, CaffeNet, VGG-M, and VGG-S have exhibited considerable function with accuracy of 0.9407, 0.9435, 0.9437, 0.9443, 0.9448, and 0.946 correspondingly. However, the implied QAFSOMN-RF and QAFSOMN-GBT schemes have attained a maximum classification accuracy of 0.9784 and 0.9871.

Table 1 Accuracy Analysis of Proposed Methodswith Various CNN Models

Methods	**Accuracy**
AlexNet	0.9437
CaffeNet	0.9443
VGG-F	0.9435
VGG-M	0.9448
VGG-S	0.9460
VGG-VD16	0.9407
VGG-VD19	0.9315
PlacesNet	0.9144

Proposed QAFSOMN-RF	**0.9784**
Proposed QAFSOMN-GBT	**0.9871**

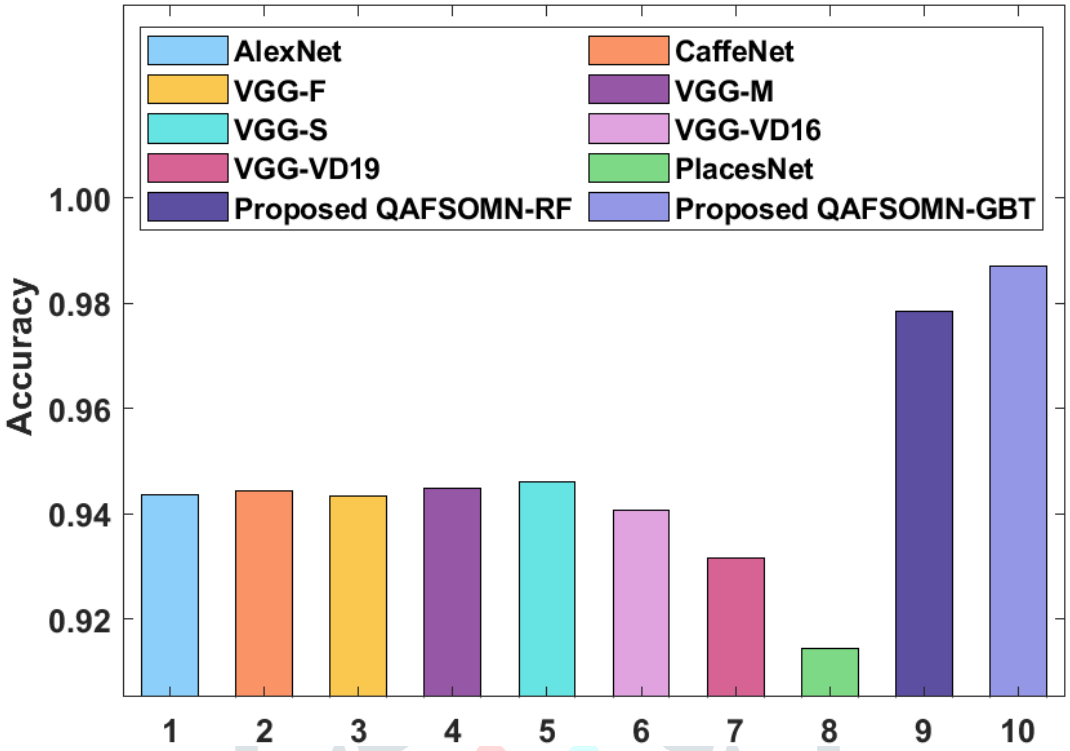

Fig. 6. Accuracy analysis of proposed model with different CNN models

In order to validate the classification results of the projected method, an extensive comparative study has been developed with traditional CNN related approaches by means of precision, recall, and F-score, as depicted in Table 2 and Fig. 7 [19]. The table measures have pointed that the CA-ResNet-BiLSTM technique has exhibited ineffective performance by gaining minimum precision of 0.7794, recall of 0.8902, and F-score of 0.8147. Followed by, the VGG-RBFNN scheme has managed to show moderate results by accomplishing a precision of 0.7818, recall of 0.8391, and F-score of 0.788. Addition, the CA-GoogLeNet-LSTM scheme has represented considerable outcomes over the former approaches by obtaining a precision of 0.7852, recall of 0.886, and F-score of 0.8178. In line with this, the VGGNet technology has depicted reasonable function with the precision of 0.7906, recall of 0.823, and F-score of 0.7854. Similarly, the CA-VGG-BiLSTM algorithm has achieved acceptable results with the precision of 0.7933, recall of 0.8399, and F-score of 0.7978 while maximum precision of 0.799, recall of 0.8614, and F-score of 0.7978 has been accomplished by the CA-ResNet-LSTM model.

Table 2 Performance Analysis of Proposed Methodswith Various Models in terms of Precision, Recall, F-score

Model	**F-score**	**Precision**	**Recall**
VGGNet	0.7854	0.7906	0.8230
VGG-RBFNN	0.7880	0.7818	0.8391
CA-VGG-LSTM	0.7957	0.8064	0.8247
CA-VGG-BiLSTM	0.7978	0.7933	0.8399

GoogLeNet	0.8068	0.8051	0.8427
GoogLeNet-RBFNN	0.8154	0.7995	0.8675
CA-GoogLeNet-LSTM	0.8178	0.7852	0.8860
CA-GoogLeNet-BiLSTM	0.8182	0.7991	0.8706
ResNet-50	0.7968	0.8086	0.8195
ResNet-RBFNN	0.8058	0.7992	0.8459
CA-ResNet-LSTM	0.8136	0.7990	0.8614
CA-ResNet-BiLSTM	0.8147	0.7794	0.8902
Proposed QAFSOMN-RF	**0.9554**	**0.9791**	**0.9453**
Proposed QAFSOMN-GBT	**0.9673**	**0.9896**	**0.9524**

In addition, the CA-GoogLeNet-BiLSTM scheme has demonstrated reasonable classification results with the precision of 0.7991, recall of 0.8459, and F-score of 0.8182. At the same time, the ResNet-RBFNN approach has managed to illustrate moderate function and resulted the precision of 0.7992, recall of 0.8459, and F-score of 0.8058. Meantime, the GoogLeNet-RBFNN technology has generated better performance with the precision of 0.7995, recall of 0.8675, and F-score of 0.8154. Moreover, the GoogLeNet scheme has gained acceptable precision of 0.8051, recall of 0.8427, and F-score of 0.8068 while closer optimal classification results are demonstrated by the CA-VGG-LSTM model with the precision of 0.8064, recall of 0.8247, and F-score of 0.7957. Even though the similar classification outcome is depicted by the ResNet technique with the precision of 0.8086, recall of 0.8195, and F-score of 0.7968, the projected QAFSOMN-GBTand QAFSOMN-RF technologies haveshowcased supreme performance with the precision of 0.9673, 0.9554 recall of 0.9524, 0.9453, and F-score of 0.9673, 0.9554.

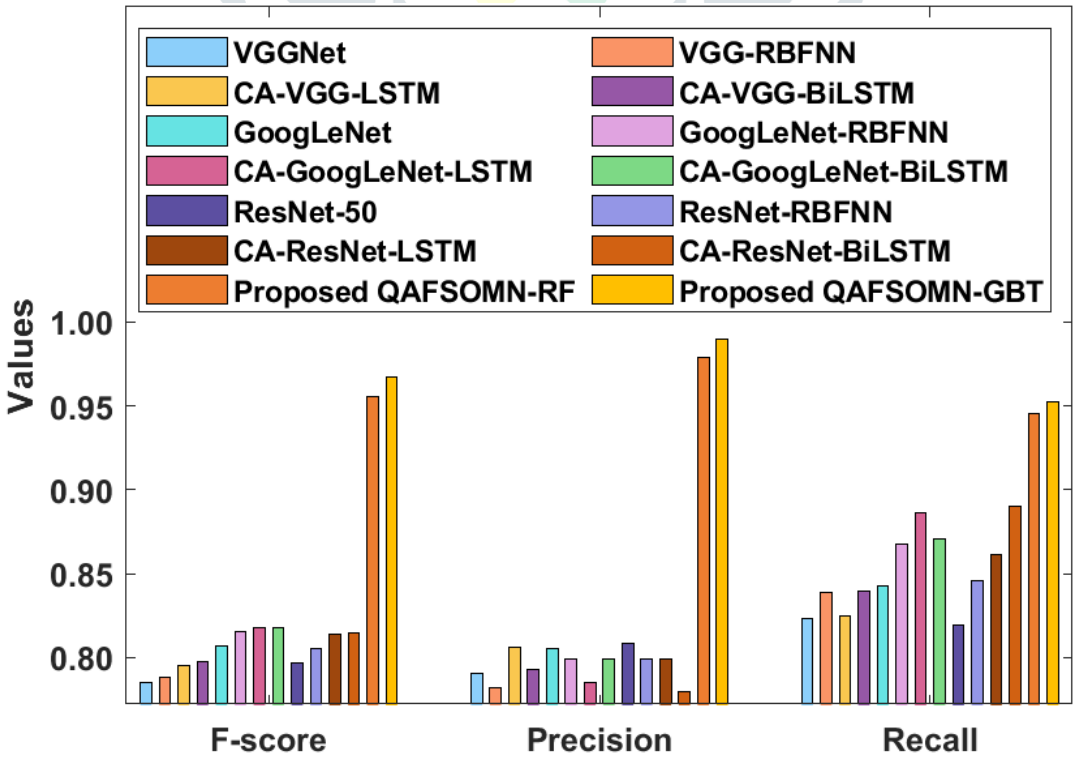

Fig. 7. Result analysis of proposed model

Table 3 and Fig. 8 referred the comparative results analysis of the ADRN-LR method with classical methods [20]. The final values have depicted the poor classification of the SCK approach with minimal accuracy of 0.7252. Moreover, the SPM technique has generated considerable accuracy of 0.7400, which exceeds the function of the SCK. Meanwhile, the SPCK++ and SC + Pooling have exhibited closer performance with the accuracy of 0.7738 and 0.8167 correspondingly. In line with this, the SG+UFL and CCM-BOVW schemes have illustrated better results with similar accuracy of 0.8664. Likewise, the PSR, UFL-SC, MSIFT, and COPD methodologies supreme results with accuracy of 0.8910, 0.9026, 0.9097, and 0.9133.

Table 3 Accuracy Analysis of Proposed Methodswith Various State of Art Methods

Methods	**Accuracy**
SPM	0.7400
SCK	0.7252
SPCK++	0.7738
SC+Pooling	0.8167
SG+UFL	0.8664
CCM-BOVW	0.8664
PSR	0.8910
UFL-SC	0.9026
MSIFT	0.9097
COPD	0.9133
Dirichlet	0.9280
VLAT	0.9430
OverFeat	0.9091
GoogleNet+Fine Tune	0.9710
CCP-Net	0.9752
Proposed QAFSOMN-RF	**0.9784**
Proposed QAFSOMN-GBT	**0.9871**

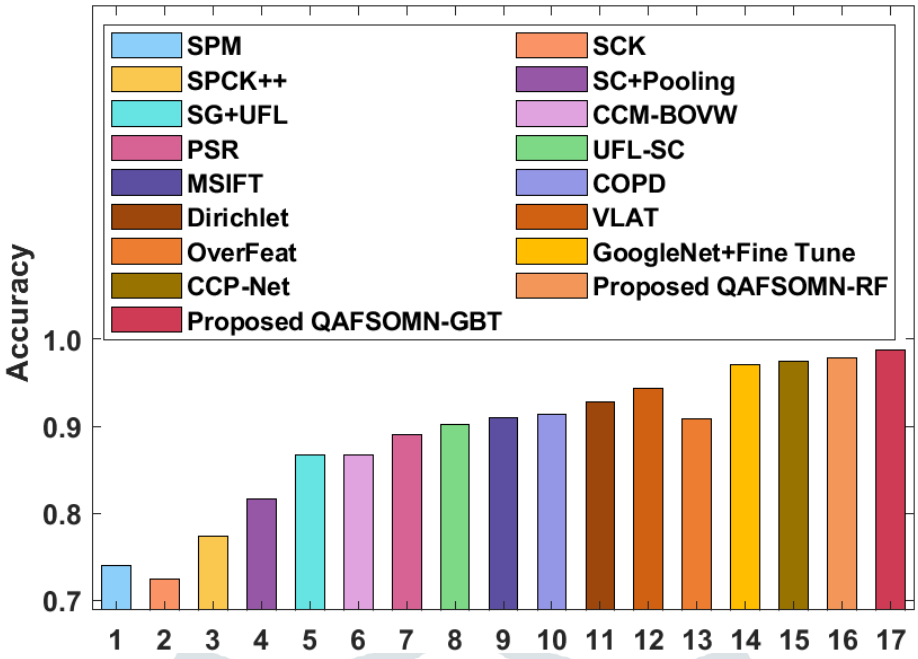

Fig. 8. Accuracy analysis of proposed model with traditional models

Followed by, the Dirichlet, VLAT, and Over Feat technologies have referred that moderate results with accuracy of 0.9280, 0.9430, and 0.9091 correspondingly. In addition, the GoogleNet + Fine Tune and CCP-Net methods have implied acceptable results with accuracy of 0.971 and 0.9752. However, the proposed QAFSOMN-RF and QAFSOMN-GBT technologies haveattained maximum classification accuracy of 0.9784 and 0.9871.

Table 4 and Fig. 9 demonstrated a running time analysis of the projected model is considered with traditional CNN methods on the given test dataset [21]. The gained values implied that the VGG-VD19 scheme requires massive running time with 496s while the VGG-VG16 approach has represented moderate performance by requiring a running time of 423s. Besides, it is marked that the VGG-M and VGG-S methods have required a considerable running time of 124s and 135s correspondingly. Along with that, the CaffeNet and AlexNet schemes have implemented closer running time of 85s and 86s. Even though the VGG-F and Places Net techniques have attained acceptable and similar running time of 82s, the presented QAFSOMN-RF and QAFSOMN-GBTmethodologies have required only a minimum running time of 68s and 66s.

Table 4 Running Time (s) Analysis of Proposed Methods with various CNN models

Methods	**Running Time (s)**
AlexNet	086
CaffeNet	085
VGG-F	082
VGG-M	124
VGG-S	135
VGG-VD16	423
VGG-VD19	496
PlacesNet	082
Proposed QAFSOMN-RF	**068**

Proposed QAFSOMN-GBT	**066**

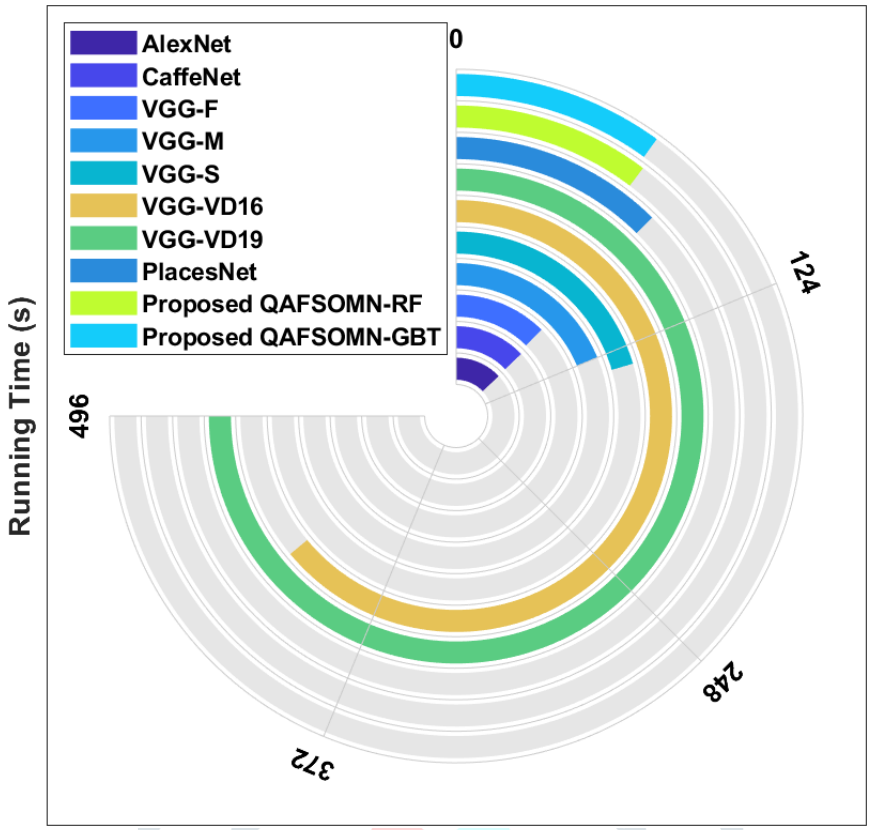

Fig. 9. Running time analysis of proposed model

From the pre-defined tables and figures, it is pointed that the proposedmethod is assumed as efficient remote sensing image scene classification process. The projected proposedapproach has attained best classification function with lower running time under various test images.

5. Conclusion

This study has presented a novel DL based QAFSOMN model for aerial image classification.Primarily, the input aerial image is preprocessed to boost up the quality of the image. Afterward, QAFSOMN model is applied as a feature extraction technique to derive an actual set of feature vectors. At last, the feature vectors are fed to the GBT and RF models to perform classification task.A set of simulations were carried out to ensure the superior classification performance on the remote sensing images.The utilization of the QAFSO algorithm paves a way of achieving higher detection performance compared to the existing methods. On the detailed experimental validation process on remote sensing imagery dataset, the QAFSOMN-GBT and QAFSOMN-RF models have obtained a maximum accuracy of 0.9871 and 0.9784. Therefore, it can be employed as an effective tool for remote sensing image classification in the real time scenarios.

References

[1] A. Darwish, A. E. Hassanien, and S. Das, ‘‘A survey of swarm and evolutionary computing approaches for deep learning,’’ Artif. Intell. Rev., vol. 53, no. 3, pp. 1767–1812, Mar. 2020.

[2] R. K. Dewangan, A. Shukla, and W. W. Godfrey, ‘‘Three dimensional path planning using grey wolf optimizer for UAVs,’’ Int. J. Speech Technol., vol. 49, no. 6, pp. 2201–2217, Jun. 2019.

[3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ‘‘Imagenet classification with deep convolutional neural networks,’’ in Proc. Adv. Neural Inf. Process. Syst. (NIPS), 2012, pp. 1097–1105.

[4] Y. Yu, Y. Yuan, H. Guan, D. Li, and T. Gu, ‘‘Aeroplane detection from high-resolution remotely sensed imagery using bag-of-visual-words based Hough forests,’’ Int. J. Remote Sens., vol. 41, no. 1, pp. 114–131, 2020.

[5] F. Al-Turjman, H. Zahmatkesh, and L. Mostarda, ‘‘Quantifying uncertainty in Internet of medical things and big-data services using intelligence and deep learning,’’ IEEE Access, vol. 7, pp. 115749–115759, 2019.

[6] L. Ye, L. Wang, Y. Sun, R. Zhu, and Y. Wei, ‘‘Aerial scene classification via an ensemble extreme learning machine classifier based on discriminative hybrid convolutional neural networks features,’’ Int. J. Remote Sens., vol. 40, no. 7, pp. 2759–2783, Apr. 2019.

[7] L.-J. Zhao, P. Tang, and L.-Z. Huo, ‘‘Land-use scene classification using a concentric circle-structured multiscale Bag-of-Visual-Words model,’’ IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 7, no. 12, pp. 4620–4631, Dec. 2014.

[8] G. Cheng, J. Han, P. Zhou, and L. Guo, ‘‘Multi-class geospatial object detection and geographic image classification based on collection of part detectors,’’ ISPRS J. Photogramm. Remote Sens., vol. 98, pp. 119–132, Dec. 2014.

[9] A. M. Cheriyadat, ‘‘Unsupervised feature learning for aerial scene classification,’’ IEEE Trans. Geosci. Remote Sens., vol. 52, no. 1, pp. 439–451, Jan. 2014.

[10] F. Zhang, B. Du, and L. Zhang, ‘‘Saliency-guided unsupervised feature learning for scene classification,’’ IEEE Trans. Geosci. Remote Sens., vol. 53, no. 4, pp. 2175–2184, Apr. 2015.

[11] A. Coates, A. Ng, and H. Lee, ‘‘An analysis of single-layer networks in unsupervised feature learning,’’ in Proc. 14th Int. Conf. Artif. Intell. Statist., 2011, pp. 215–223.

[12] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, ‘‘Gradient-based learning applied to document recognition,’’ Proc. IEEE, vol. 86, no. 11, pp. 2278–2324, Nov. 1998, doi: 10.1109/5.726791.

[13] M. Cimpoi, S. Maji, and A. Vedaldi, ‘‘Deep filter banks for texture recognition and segmentation,’’ in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2015, pp. 3828–3836.

[14] M. Castelluccio, G. Poggi, C. Sansone, and L. Verdoliva, ‘‘Land use classification in remote sensing images by convolutional neural networks,’’ 2015, arXiv:1508.00092. [Online]. Available: http://arxiv. org/abs/1508.00092

[15] O. A. B. Penatti, K. Nogueira, and J. A. dos Santos, ‘‘Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?’’ in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), Jun. 2015, pp. 44–51.

[16] Rahnamayan, S., Tizhoosh, H.R. and Salama, M.M., 2007, September. Quasi-oppositional differential evolution. In *2007 IEEE congress on evolutionary computation* (pp. 2229-2236). IEEE.

[17] http://weegee.vision.ucmerced.edu/datasets/landuse.html

[18] Rajagopal, A., Joshi, G.P., Ramachandran, A., Subhalakshmi, R.T., Khari, M., Jha, S., Shankar, K. and You, J., 2020. A Deep Learning Model Based on Multi-Objective Particle Swarm Optimization for Scene Classification in Unmanned Aerial Vehicles. IEEE Access, 8, pp.135383-135393.

[19] Hua, Y., Mou, L. and Zhu, X.X., 2019. Recurrently exploring class-wise attention in a hybrid convolutional and bidirectional LSTM network for multi-label aerial image classification. ISPRS journal of photogrammetry and remote sensing, 149, pp.188-199.

[20] Qi, K., Guan, Q., Yang, C., Peng, F., Shen, S. and Wu, H., 2018. Concentric circle pooling in deep convolutional networks for remote sensing scene classification. Remote Sensing, 10(6), p.934.

[21] Rajagopal, Aghila, A. Ramachandran, K. Shankar, Manju Khari, Sudan Jha, Yongju Lee, and Gyanendra Prasad Joshi. "Fine-tuned residual network-based features with latent variable support vector machine-based optimal scene classification model for unmanned aerial vehicles." IEEE Access 8 (2020): 118396-118404.