Aircraft Classification Based on PCA and Feature Fusion Techniques in Convolutional Neural Network

Article *in* IEEE Access · December 2021

DOI: 10.1109/ACCESS.2021.3132062

6 authors, including:

Wazir Zada Khan
University of Wah
136 PUBLICATIONS **6,239** CITATIONS
SEE PROFILE

Mohammed Y Aalsalem
Jazan University
74 PUBLICATIONS **2,181** CITATIONS
SEE PROFILE

Heejung Yu
Korea University
170 PUBLICATIONS **2,698** CITATIONS
SEE PROFILE

Yousaf Bin Zikria
TIIS Sydney
133 PUBLICATIONS **5,392** CITATIONS
SEE PROFILE

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received November 22, 2021, accepted November 30, 2021, date of publication December 1, 2021,
date of current version December 14, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3132062

Aircraft Classification Based on PCA and Feature Fusion Techniques in Convolutional Neural Network

FAISAL AZAM[1], AKASH RIZVI[1], WAZIR ZADA KHAN[2], (Senior Member, IEEE), MOHAMMED Y. AALSALEM[3], HEEJUNG YU[4], (Senior Member, IEEE), AND YOUSAF BIN ZIKRIA[5], (Senior Member, IEEE)

[1]Department of Computer Science, COMSATS University Islamabad, Wah Campus, Rawalpindi 47040, Pakistan
[2]Department of Computer Science, University of Wah, Wah 47040, Pakistan
[3]Department of Computer Science and Information Technology, Jazan University, Jizan 45142, Saudi Arabia
[4]Department of Electronics and Information Engineering, Korea University, Sejong 30019, South Korea
[5]Department of Information and Communication Engineering, Yeungnam University, Gyeongsan 38541, South Korea

Corresponding authors: Heejung Yu (heejungyu@korea.ac.kr) and Yousaf Bin Zikria (yousafbinzikria@ynu.ac.kr)

This work was supported in part by the National Research Foundation of Korea (NRF) Grant through the Ministry of Science and ICT (MSIT), South Korea, under Grant 2019R1A2C1083988; in part by the Basic Science Research Program through the NRF funded by the Ministry of Education under Grant 2021R1I1A3041887; and in part by the MSIT, South Korea, through the Information Technology Research Center (ITRC) Support Program supervised by the Institute for Information and Communications Technology Promotion (IITP) under Grant IITP-2021-2016-0-00313.

ABSTRACT The characterization of aircraft in remote sensing satellite imagery has many armed and civil applications. For civil purposes, such as in tragedy and emergency aircraft searching, airport scrutiny and aircraft identification from satellite images are very important. This study presents an automated methodology based on handcrafted and deep convolutional neural network (DCNN) features. The presented aircraft classification technique consists of the following steps. The handcrafted features achieved from a local binary pattern (LBP) and DCNN are fused by feature fusion techniques. The DCNN features are extracted from Alexnet and Inception V3. Then we adopted a feature selection technique called principal component analysis (PCA). PCA removes the redundant and irrelevant information and improves the classification performance. Then, Famous supervised methodologies categorize these selected features. We chose the best classifier based on its highest accuracy. The proposed technique is executed on the multi-type aircraft remote sensing images (MTARSI) dataset, and the overall highest accuracy that we achieved from our proposed method is 96.8% by the linear support vector machine (SVM) classifier.

INDEX TERMS Aircraft classification, CNN, feature extraction, feature fusion, identification of aircraft.

I. INTRODUCTION

In public and martial applications, recognition of aircraft type from remotely sensed imageries has more importance. In this era, images can be found with high spatial resolution remote sensing by using modern technologies and equipment. With the progress in remote sensing technologies, the detail attributes of a target can be obtained due to the enhanced resolution. Characterization and identification of aircraft have accomplished research and investigational attention. It has a prodigious consequence in aerospace fields, applications, intellect evidence, and much more [1]. For civil purposes, such as emergency aircraft searching, identification of an aircraft, and airport scrutiny are extremely important [2], [3].

In the early stages of researches, handcrafted features, like "SIFT" [4], [5] and "HOG" [6], are some of the approaches that were used for the recognition of objects from remote sensing images such as aircraft, boats, houses and so on. Numerous methodologies are based on shape matching methods [7], [8], like the grouping of an edge potential and artificial bee colony (ABC) methodology in [8] and the coarse-to-fine, suggested in [7] by employing the parametric shape representation. These technologies play a key part in the presentation improvements of aircraft recognition/acknowledgment. With the expansion and advancement of hardware efficiency, deep neural networks have

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

revolutionized remote sensing satellite images. Deep convolutional neural network (DCNN) plays an important role and has been broadly applied in different fields such as segmentation [9], [10], identification or detection [11], [12], cataloging [13], [14], etc.

Feature extraction is also a vigorous chapter for all computerized systems. The features merging and selection procedures present much devotion last couple of years in computer vision (CV), and various associated techniques are presented, expanding the system recognition correctness [15]–[17]. The synthesis of several features gives improved performance compared to a particular feature kind. The noteworthy advantage of feature fusion is to associate the facts of multiple descriptions, which enhances the complete system efficiency. The drawback of feature fusion is to upsurge the recognition period due to the accumulation of redundant and unrelated evidence. These types of issues are fixed by the features selection phases, which eliminates the redundant and unrelated evidence and only picks the top features [1], [18]. Currently, deep learning (DL) illustrates more achievements in machine learning (ML) and CV research fields, particularly in surveillance jobs, classification [19], biometrics [20], satellite imageries [21], medical imaging [22]. In a convolutional neural network (CNN), the extractive features enclose both regional and global information.

Image recognition is the procedure of identification and acknowledgment of an element in digital pictures or videos. Object recognition in digital imageries would possibly start with pre-processing image procedures, for example, image enhancement, noise removal, that are treated by feature extraction to discover sections, lines, and possible zones with specific exteriors. Besides the composite structure, altered aircraft vary in figures, shading, scope, or colors, and even for one part of the airplane. The intensity and texture are typically dissimilar in various situations. Furthermore, recognition frequently suffers from several instabilities, for example, altered contrasts, cluttering, and anxiety inconsistency. Subsequently, the resistance to disruption and robustness are highly required for the methodology.

Several approaches were applied on different datasets under different investigational situations. However, the datasets used are often not publicly available. That is why it is very challenging to reproduce the effort for comparison. A dataset called multi-type aircraft remote sensing images (MTARSI) [14] is now publicly available to solve this problem. There are 20 airplane types with 9,385 imageries with amalgamated backgrounds and dissimilar three-dimensional (3-D) resolutions. In our investigation, after applying some prior processing on aircraft images, we apply handcrafted and some convolutional neural network algorithms on MTARSI dataset to improve classification accuracy. The global average pooling layer (APL) was utilized to figure the average of every feature plot from the preceding layer. We reduce the redundant information and irrelevant features by using the principal component analysis (PCA) technique [23], [24]. After that, we perform the feature fusion technique. This technique associates the facts of multiple imageries and enhances the complete system efficiency. Some of the other classification methods that we applied are support vector machine (SVM) (Linear and Quadratic) [25], least squares SVM (LSSVM) [26], and k-nearest neighbors (KNN) [27].

There are various challenges for object cataloging in aerial images, which vitiates the system's accuracy. These challenges can be the similarity between multiple objects, illumination effects, resolution of aerial images, complex and transparent background. Several methodologies are presented in the literature, but there is scope to handle these types of challenges. The size of the dataset is also challenging in aircraft classification for the training of the models. Our work used the MTARSI dataset, which contains 20 classes of aircraft and ranges from 230∼800 images per class. We present a methodology for aircraft classification by using deep learning techniques. The major contribution in our work is listed below:

- Local binary pattern (LBP) features are computed for texture information of aircrafts
- Feature extraction by CNN along with LBP
- Feature selection by using PCA
- After selecting the features, different classifiers are applied, and the best result is compared with existing techniques.

II. RELATED WORK

In this part of our work, we will discuss several datasets commonly used for aircraft recognition. In public and martial applications, recognition of aircraft from remotely sensed images has more significance. On the other hand, the datasets used are often not publicly available. That is why it is very challenging to reproduce the effort for comparison. At present, there are five popular datasets for aircraft identification named, University of California Merced land use (UCMerced_LandUse) [10], Pattern-Net [28], North Western Polytechnical University-remote sensing image scene classification (NWPU-RESISC-45) [9], fine-grained visual classification of aircraft (FGVC-Aircraft) [29] and MTARSI [14].

The UCMerced_LandUse [10] was established at the University of California, Merced. It is a widely used dataset in the area of satellite images, especially for object classification. The images were primarily selected from the US geological survey and then randomly cropped into 256 × 256 pixels. The spatial resolution is ∼0.3m. The dataset consists of 21 different classes i.e., Airplane, Baseball, runway, residential, storage tanks, chaparral, forest, etc. The dataset contains 2100 images, 100 per class.

In 2017, a dataset, called NWPU-RESISC-45 [9], was created by North-western Polytechnical University, which is freely available as a benchmark for remotely sensed images scene classification. The RESISC-45 dataset has 31,500 images consist of 45 classes which contain 700 imageries in a distinct category.

The Pattern-Net [28] dataset was created in the laboratory of Wuhan University, China, and the University of California,

USA. The dataset is the largest high-resolution Satellite images dataset. It has 30,400 images with 38 different categories such as baseball, airplanes, beach, cemeteries, shrubs, bridges, swimming pools, tanks, tennis courts, etc. Each class contains 800 images having spatial resolutions from 0.062m $\sim$ 4.693m and 256 $\times$ 256 sizes.

FGVC-Aircraft [29] is a dataset for the visual classification of aircraft. It contains 10,200 images of 102 different aircraft's model variants 100 images for each class. The airplane in every image is marked with a bounding box and an ordered aircraft model label. In all of above mentioned datasets, i.e., UCMerced_LandUse [10], NWPU-RESISC-45 [9], PatternNet [28], and FGVC-Aircraft [29] might be used to train the airplane recognition, identification, and segmentation algorithms. However, in these, the airplanes are used as a sub-category in the dataset. Now a dataset called MTARSI complied by Wu, Z.-Z., *et al.* [14] is now publicly available, which consists of a wide variety of aircraft. There are 9,385 imageries of 20 aircraft with an amalgamated environment and distinct space resolutions. In MTARSI [14], different methodologies have been applied to identify the aircraft type. The detail of the previously presented remote sensing satellite dataset is shown in **Table 1**.

Many other pieces of research have been done on aerial images of aircraft, as in multiple class activation mapping (MultiCAM), which was used to pull out the diverse portions of aircraft of several styles [1]. Identification of aircraft was based on corner clustering, and CNN was proposed in [30]. Detecting small objects like aircraft from remotely sensed imagery using YOLOv3 achieved noteworthy detection performance with a small processing overhead [31]. Recognition of boats and airplanes in long-distance images by the composition of deep features attained from CNN [32]. Li *et al.* implemented an efficient aircraft identification agenda based on a reinforcement learning and CNN (RL-CNN) model in [33]. This method was used to correctly and quickly locate the airplanes in long distanced images.

TABLE 1. Publicly available remote sensing dataset.

Datasets	Scene Class	Image / class	Total Images	Image sizes	Spatial Resolution	Year
UC Merced [10]	21	100	2,100	256 x 256	0.3	2010
WHU-RS19 [34]	19	~50	1,005	600 x 600	0.5	2012
SIRI-WHU [35]	12	200	2,400	200 x 200	2	2016
RSSCN7 [36]	7	400	2,800	400 x 400	--	2015
RSCI1 [37]	11	~100	1232	512 x 512	0.2	2016
Brazilian Coffee scene [38]	2	1438	2876	64 x 64	--	2015
NWPU-RESISC45 [9]	45	700	31,500	256 x 256	~0.3 – 0.2	2016
PatternNet [28]	38	800	30,400	256 x 256	4.69-0.06	2018
MTARSI [14]	20	230-846	9,385	--	0.3 – 1.0	2020

III. PROPOSED METHODOLOGY

Our proposed methodology estimates the overall identification procedures and neural network tactics for airplane type identification on the MTARSI data package. It consists of the following three-step procedure: classical and CNN feature extraction, selection, and fusion of best-selected features. We apply PCA to the aerial images of aircraft after transferring them to the pre-trained DCNN models. PCA improves the classification performance. After that, we perform feature fusion on the dataset. We conduct a series of experiments with CNN, i.e., VGG16, AlexNet, Resnet, and inception. The flow diagram of our suggested model is shown in **Figure 1**. Our diagram indicates that the CNN and classical features are extracted from our dataset in parallel processing and choose the best features before the fusion stage. Lastly, we apply classifiers to our dataset to get the labeled images of aircraft.

As revealed in **Figure 1**, the input images are pre-processed and passed to the handcrafted feature LBP. Simultaneously, the dataset is given to CNN such as Alexnet and Resnet for feature extractions. Afterward, from these extracted features, we select the most robust feature by using PCA. After that, we performed feature fusion methodology on the obtained finest subset features. Finally, we perform different classifiers, i.e., SVM, KNN, ESD, for final identification. A comprehensive explanation of each phase is provided below sections.

A. FEATURE EXTRACTION

Feature extraction is one of the main procedures in computer vision to demonstrate an object in the picture. The working of any automated technique depends on the number of extracted features. The robust and related features give improved accuracy, but the noisy or redundant features vitiate the system outcomes. We computed the LBP technique in the classical feature. Whereas the pre-trained model named Alexnet, inception V3 is utilized in CNN. The comprehensive description of these features explains below.

1) LBP

We extract LBP features from the greyscale images to handle the complications of illumination changes and simplify the complexity of originally extracted LBP features. **Figure 2** depicts the working of LBP.

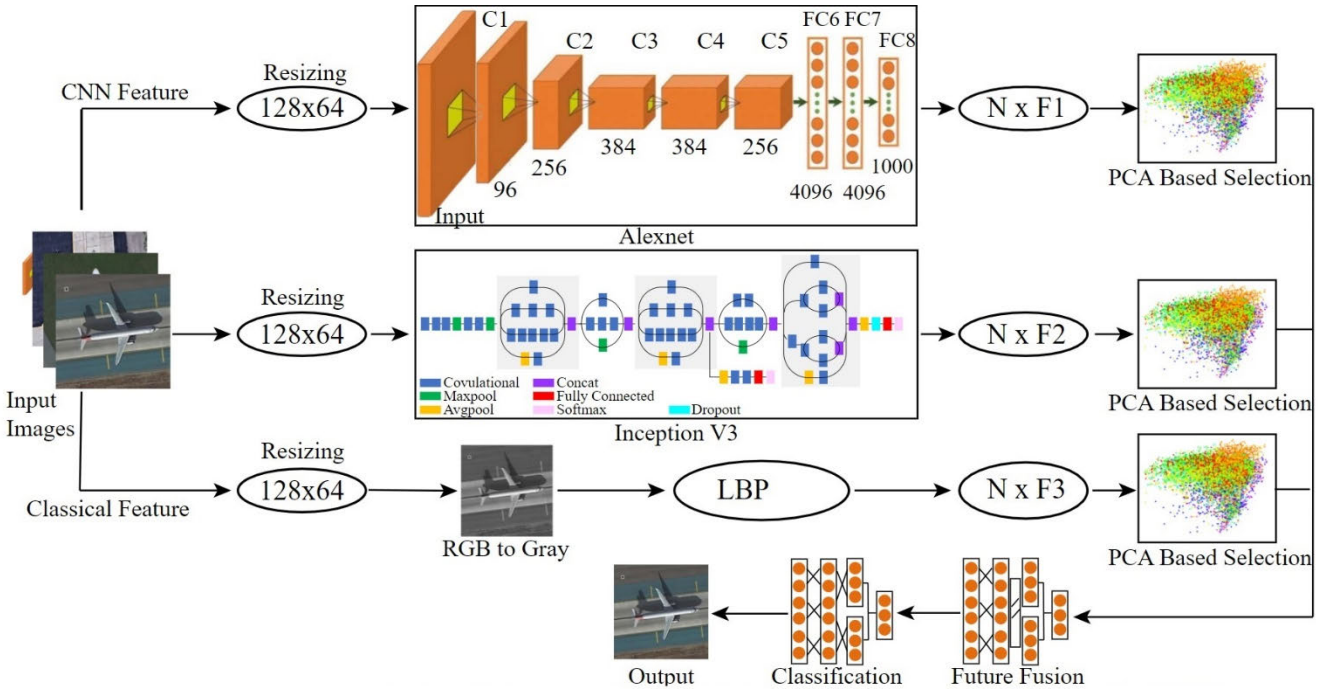


FIGURE 1. Flow architecture of the proposed CNN and future fusion based remotely sensed image classification.

12	15	18
5	8	3
8	1	2

Original Pixel Value

1	1	1
0		0
1	0	0

Binary Pixel Value

Binary number will be	
≥8 = 1	<8 = 0

FIGURE 2. Central pixels comparison in LBP.

It labels the pixels of an image by thresholding the neighborhood of each pixel and considers the result as a binary number. The central pixel is compared with each neighbor pixel and assigned a binary 1 or 0. If the value of the pixel is less than the central pixel, then the value of that pixel will be binary 1; otherwise, 0.

The achieved binary code can be written from the topmost first cell and moving to the right from the above figure, and the binary code will be.

1	1	1	0	0	0	1	0

Moreover, the light changes the pixel value of the image, but it does not change the binary pattern of a texture, as shown in **Figure 3**.

2) CNN FEATURE

The leading representation in dl is that of CNNs, which is assumed in an extensive range of facets in image handling, as well as in image categorization [39], super resolution restoration [40], object detection [41], etc. In CNN, we used the pre-trained CNN model name inception v3 and alexnet. **Figure 1** depicts the two of the pre-trained models, i.e., alexnet and inception-v3. Alexnet consists of 5 convolutional layers and 3 fc layers (fully connected layers).

Whereas, Inception V3 has 316 layers and 350 connections. In these models, we apply several filters on the same layer for deep feature extraction. A CNN contains three key elements like convolution layer, pooling layer, and FC layers. Each part plays a diverse task. The working procedure of CNN is revealed in **Figure 4**.

Both of these models (Alexnet and Inception V3) are initially trained on a database of ImageNet [42]. Therefore we utilized their complete architecture by applying transfer learning notion and executing training on MTARSI dataset. We divide the dataset into 70:30 ratios for training and testing purposes. Then train Alexnet and Inception V3 on MTARSI dataset by utilizing transfer learning. Our proposed reduction, Traditional and DCNN feature fusion base model, is revealed in **Figure 1**.

In CNN, It is a very suitable way to extract automatically extreme connected features [43]. It gets input as $X \times H \times 3$ dimensional matrix. The thresholding of the layer is t, and size of the Kernel (K) is associated with the convolutional layer as $x \times h \times 3$. The chief formulation of the convolutional

12	15	18
5	8	3
8	1	2

1	1	1
0		0
1	0	0

Binary Number Will	
$\geq 8 = 1$	$\leq 8 = 0$

1	1	1	0	0	0	1	0

FIGURE 3. Light effect on binary pattern.

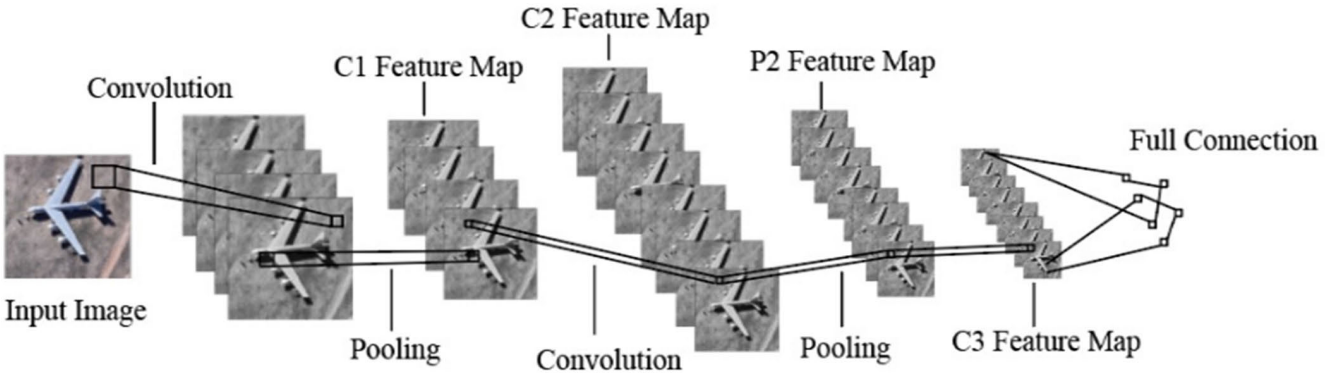


FIGURE 4. Working procedure of CNN model.

layer is well defined by the mathematical Equation 1-3.

$$G_n = \frac{G - g + 2 \times Z}{D} + 1 \tag{1}$$

$$K_n = X \times K = \sum_{i=1}^{3} (x_i + k_i) + t \tag{2}$$

$$K_n = \frac{K - k + 2 \times Z}{D} + 1 \tag{3}$$

where D denotes the number of convolutions, K represents the kernel size, and t represents the threshold value. Afterward, ReLu activation layer [44] is executed as follows:

$$y = \max(0, X) \tag{4}$$

After that, one more layer, called pooling, is executed to diminish the dimensionality of the extracted attributes from the preceding layers. There are three sorts of pooling layers that are commonly used, i.e., maximum, minimum, and average [45]. The benefit of the pooling procedure is that we can attain comparative features. **Figure 5** depicts the pooling procedure.

Lastly, the final layer of CNN architecture is the fully connected (FC) layer; its mathematical representation follows.

$$\mathrm{L^{in}} = \mathrm{X} \times \mathrm{W} + \mathrm{D} \tag{5}$$

$$\mathrm{L^{out}} = \mathrm{ReLU}(\mathrm{L^{in}}) \tag{6}$$

$\mathrm{L^{in}}$ stands for input layer, where $\mathrm{L^{out}}$ stands for the output layer. The input layer ($\mathrm{L^{in}}$) is passed to the Relu activation function, and then the resultant layer is represented as $\mathrm{L^{out}}$.

The generalized equation of FC layer is as follows:

$$L^{out} = X \tag{7}$$

$$L_i^{in} = L_{i-1}^{out} \times W + D_i \tag{8}$$

$$L_i^{out} = F_i(L_i^{in}) \tag{9}$$

where F_i denotes the activation function at the layer i.

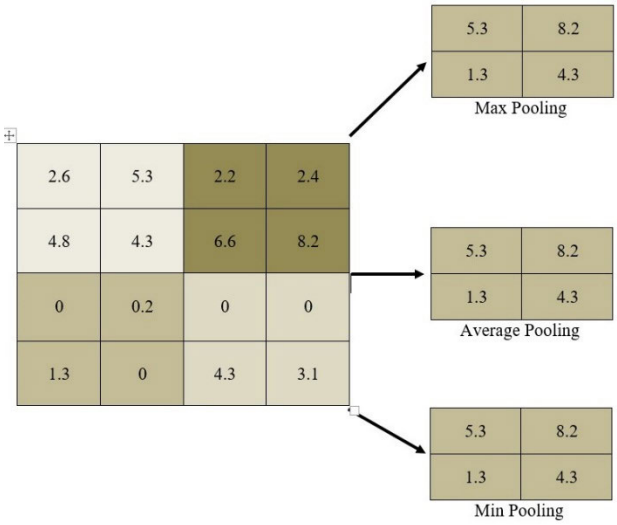


FIGURE 5. Example of pooling procedure.

B. FEATURE SELECTION

In AI and ML, feature selection is the method of attaining the minimum number of robust features from an innovative set with the least data loss. The researchers try hard to seek various methods to eliminate the glitches of massive amounts of data into little portions. The higher dimensional feature increases the computational cost, memory of algorithm, and accommodation. Therefore, an algorithm requires that it is effective enough to remove the redundant information. This algorithm may also handle the irrelevant feature. We used a selection method that removes the unrelated feature and reduces the unnecessary information in our work. **Figure 1** shows the comprehensive feature extraction and selection process. The notation ***F1, F2*** and ***F3*** shows the extracted feature from Alexnet, Inception V3, and LBP. The notation ***N*** represents the entire number of images used for testing and training purposes. The features extracted from LBP, Alexnet, and Inception V3 are then passed to PCA based selection method. These selected features are then fused and carried out further classification. The detailed working of PCA is described below.

PCA is numerically a difficult procedure to accomplish this overview. The technique generates a novel set of variables called principal components (PCs). The entire PCs are orthogonal to each other, so there is no superfluous information. PCA is a methodology that takes numeric datasets and utilizes orthogonal techniques of transformation. It transforms an inspection into a variable set, then plotted with a set of variables recognized as PCs. When there are noisy data sets, PCA is most beneficial as it is much easier if the inconsistency spreads on some of the components instead of over the entire set. Thus relatively, there is less noise effect as the signal-to-noise-ratio of the initial higher some components. This consequence of focusing much of the sign on the initial few components can be attained by PCA's dimensionality reduction attributes. Later, PCs may be conquered by noise, and consequently, they can be rejected without immense loss. In addition, this method reduces the dimension of the dataset. However, the variance of the dataset remains the same. Feature selection diminishes the merged feature vector (FV) and chooses the most favorable features for well recognition. Principal components as a whole provide an orthogonal foundation for the data space [23]. As a result, we offer unsupervised FS algorithms for PCA based on eigenvectors analysis to recognize the original features.

The Eigenvalue decomposition of the data covariance /correlation matrix or the singular value decomposition of a data matrix is used to determine PCs. Usually, after each attribute's data has been mean-centered. When the variances of variables are significantly high as compared to correlation, a covariance matrix is preferred. When the variables are of various kinds, it is preferable to use type correlation.

We can say that the PCA consists of these four key steps: (a) first getting the mean of fused feature Vector; (b) subtracting mean from every feature; (c) computing the covariance matrix; (d) computing the eigenvalues and vectors of the covariance matrix. The PCA returns the main components as well as a score. The algorithm of these PCA steps is given below:

Algorithm: PCA

Input: Dataset matrix [X]
Output: Features reduction
Step 1: Generate N x d dataset matrix (one row vector per data point x_n)
Step 2: subtract the mean from every vector row x_n in X.
Step 3: calculation covariance of matrix X.
Step 4: find eigenvalues and vectors of a covariance matrix.
Step 5: PCs the mean eigenvectors with the greatest eigenvalues.
Step 6: Output.

After selecting the finest subset of features with a minimum error rate and best accuracy, we further passed these selected features for fusion. The detail of the feature fusion process is given below.

C. FEATURE FUSION

Feature fusion is an energetic research area to achieve the finest accuracy compared with the individual feature sets [46] and [47]. There are two kinds of feature fusions, i) early fusion and ii) late fusion. The feature-based combination of details is "early fusion" while the late fusion is applied at the categorization step. After sampling, the superficial and deep layer attributes are combined to an identical extent to control the glitches of little dimensionality in the deep layer and the insufficient appearance of tiny stuff. The dimensions of the applicant are customized to fit the dimensions of the authentic aircraft in the aerial images.

In our presented methodology, we apply the late feature fusion technique. We provide the dataset to CNN models Alexnet, Inception V3, and a hand crafted feature LBP. After extracting the features from these models, we passed these features to PCA based selection technique. By obtaining the finest subset of features from PCA, we further fused these features. Finally, we provide these fused features to various classifiers such as Quadratic SVM (QSVM), Liner SVM (LSVM), KNN, and ensemble subspace disarmament (ESD). The best classifier was selected based on higher accuracy. The proposed experimental result is described in section V.

D. DATASET DESCRIPTION

The proposed work is evaluated on the MTARSI dataset [14], comprising several types of aircraft images. By this, the identification of airplane types from remotely sensed imageries becomes more possible. This dataset has 9,385 images of 36 different airports, including 20 types of aircraft acquired from Google Earth and manually expanded. This novel dataset of long distanced images is composes of the following 20 airplane types: A-10, A-26, B-29, B-52, B-1, B-2, Boeing, C-17, C-130, KC-10, C-5, C-135, C-21, F-22, F-16, E-3, P-63, U-2, T-43, and T-6.

Experts in the area of aerial imageries analysis cautiously label every single sample picture. Each picture includes absolutely one whole airplane. Each kind of airplane model in MTARSI data package is revealed in Figure 6. The number of model imageries of aircraft in each category is different,

FIGURE 6. Models of the 20 types of aircraft from MTARSI [14].

TABLE 2. Various aircraft types and the number of images in each class of MTARSI.

Aircraft	Images	Aircraft	Images	Aircraft	Images
C-130	763	B-1	513	A-10	345
C-135	526	B-2	619	T-6	248
C-5	499	B-52	548	P-63	305
C-17	480	B-29	321	A-26	230
E-3	452	Boeing	605	T-43	306
F-16	372	F-22	846	C-21	491
KC-10	554	U-2	362	--	--

ranging from 230 to 846. The detail of the classes and the number of images per class is shown in Table 2.

1) DATASET AUGMENTATION

The MTARSI dataset contains some models based on the differences in background, pose, resolution, light, color, and aircraft model. Some aircraft, such as the KC-10 tanker and the B-2 bomber, are very unusual and hard to capture by satellite sensors. This condition delays the procedure of accumulating and structuring the data sets. To diminish this issue, Wu *et al.* [14] preciously enlarge the dataset by pretending pictures of aircraft that were hard to witness. We randomly select the changed background from the related satellite imageries (i.e., those Lands that do not surround any airplanes. At last, the achieved extracted airplane picture is reflected, rotated, and subsequently merged with the particular background to get the ultimate resultant image. The detailed procedure is revealed in **Figure 7**.

The dataset has many variations in the images of aircraft, like the same type of airplanes in different colors, poses, points of view, changes of background, and resolution. The pictures are captured at different times, like in a day, evening, or different weather conditions, etc. The sample images are shown in **Figure 8**.

IV. EXPERIMENTAL SETUP AND RESULT

The presented CNN technique is executed in a publically available dataset named MTARSI. The dataset has 9,385 images of 36 different airports, including 20 types of aircraft. The experiment is performed on MATLAB 2018b by desktop computer Core i7 8th generation with 16GB of RAM.

The extracted CNN features are predictable by many different classifiers and picked the classifier grounded on the utmost accuracy. Different classifiers that are applied in our efforts are SVM, KNN, and ensemble approaches. All conclusions are calculated through 5-10-folds cross validation, and the 70:30 approach is utilized. Then figure the performance in the following measures, i.e., accuracy, recall, precision, F1-score.

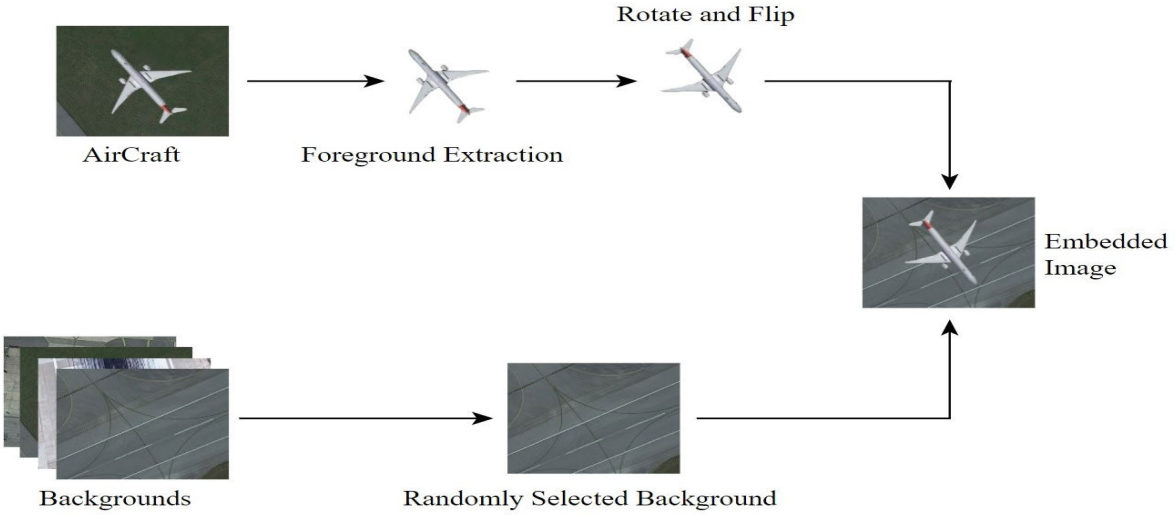


FIGURE 7. Flow diagram of remotely sensed images for image simulation procedures.

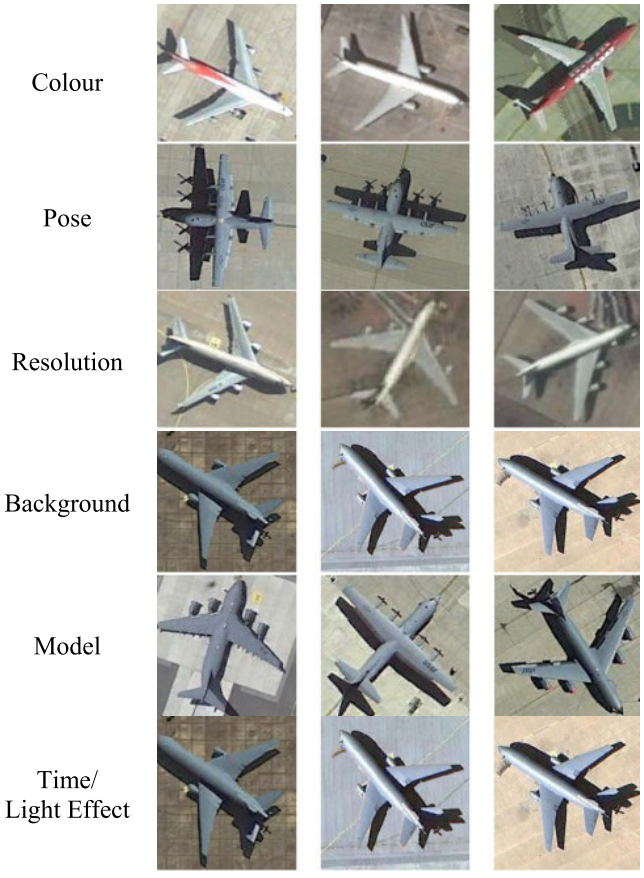


FIGURE 8. Sample images under the different condition like colour, pose, resolution background, model, time / light effect.

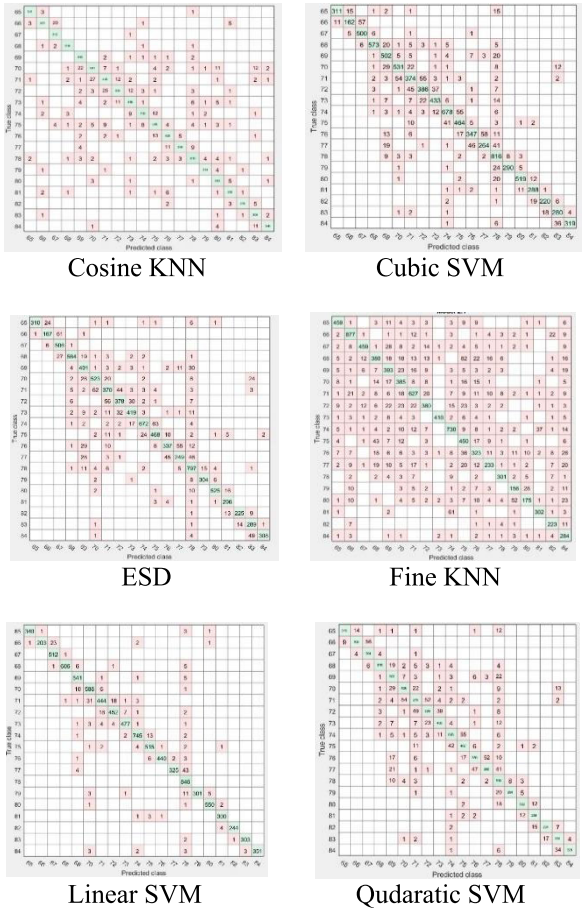


FIGURE 9. Confusion matrix of various classifiers.

A. RESULT AND ANALYSIS

In this part of our article, we presented the results of the proposed model in Tabular form. The confusion matrix is also attached. Different CNN Pre-trained models are executed in this effort and abstract features from FCs layers. After fusion, 10-fold cross-validation (10FCV) is prepared for training and

TABLE 3. Result of different classifiers obtained by our proposed model.

Classifier	Accuracy (%)	Kappa Value	Precision (%)	Recall	F1 Score	Loss
Quadratic SVM	88.1	0.88	87.43	88.77	88.09	11.9
Linear SVM	**96.8**	**0.97**	**96.34**	**97.44**	**96.89**	**3.2**
Cubic SVM	88.0	0.87	87.08	88.6	87.83	12
Cosine KNN	95.2	0.95	95.02	95.2	95.11	4.8
Fine KNN	80.4	0.79	79.34	79.67	79.50	19.6
ESD	87.4	0.87	86.8	87.7	87.26	12.6
Coarse Gaussian SVM	89.7	0.90	89.2	88.2	88.70	10.3
Medium Gaussian SVM	95.5	0.96	94.7	94.5	94.60	4.5
Medium KNN	84.4	0.84	83.8	84.6	84.20	15.6
Cubic KNN	74.7	0.75	74.21	74.33	74.27	25.3
Weighted KNN	85.9	0.86	85.03	85.5	85.26	14.1
Ensemble Boosted Tree	34.6	0.35	33.6	34.72	34.15	65.4
Ensemble Bagged Trees	69.4	0.69	68.7	68.9	68.80	30.6
Ensemble Subspace KNN	83.8	0.84	83.1	83.32	83.21	16.2
Linear Discriminant	95.9	0.96	95.02	94.2	94.61	4.1

TABLE 4. Result of different classifiers obtained by our proposed model image dimension (512 × 256).

Classifier	Accuracy (%)	Kappa Value	Precision (%)	Recall	F1 Score	Loss
Quadratic SVM	**87.5**	**0.88**	**87.15**	**89.05**	**88.09**	**12.5**
Linear SVM	86.3	0.86	85.80	88.05	86.91	13.7
Cubic SVM	**87.5**	**0.88**	**87.10**	**89.10**	**88.09**	**12.5**
Cosine KNN	83.8	0.84	83.35	83.45	83.40	16.2
Fine KNN	78.4	0.78	78.35	78.35	78.35	21.6
ESD	82.60	0.83	83.10	83.25	83.17	17.4
Coarse Gaussian SVM	73.1	0.72	69.20	82.80	75.39	26.9
Medium Gaussian SVM	86.60	0.87	85.35	88.65	86.97	13.4
Medium KNN	78.50	0.78	78.55	79.30	78.92	21.5
Cubic KNN	76.4	0.76	76.55	77.95	77.24	23.6
Weighted KNN	80.10	0.79	80.15	80.85	80.50	19.9
Ensemble Boosted Tree	29.70	0.30	27.27	35.70	30.92	70.3
Ensemble Bagged Trees	59.80	0.61	58.45	63.90	61.05	40.2
Ensemble Subspace KNN	74.00	0.73	73.70	73.85	73.77	26.0
Linear Discriminant	84	0.83	84.40	84.70	84.55	16.0
FGSVM	9.1	0.80	5.08	20.45	8.14	90.9
Coarse KNN	62.7	0.63	63.50	77.10	69.64	37.3
Coarse Tree	16.3	0.15	12.63	10.65	11.56	83.7
Medium Tree	23.4	0.22	20.98	28.42	24.14	76.6
Fine Tree	32.3	0.31	31.70	37.95	34.54	67.7
Ensemble RusBoosted Tree	30.60	0.31	34.47	31.45	32.89	69.4

testing models. The testing results are present in Table 3 and Table 4, with the maximum attained accuracy is 96.8% on the Linear SVM classifier and Precision of 96.34%, Recall of 97.44%, simultaneously. The second most achieved accuracy is of Linear Discriminant classifier that is 95.9%. The poorest achieved accuracy for feature fusion is Ensemble Boosted Tree that is 34.6%. Finally, the resultant confusion matrixes and graphs are revealed in Figure 9 and Figure 10.

1) COMPARISON

Wu *et al.* [14] carried out their simulation on MTARSI dataset to classify aircraft recognition by different state-of-the-art methodologies. In their methodology, the maximum achieved accuracy is 89.7% by the EfficientNet deep learning model, while the maximum accuracy that we achieved from our proposed methodology is 96.8%.

Zhi used the composite scaling technique for balancing the system depth, breadth, and resolution. The second finest attain accuracy is 89.6% which is calculated by applying ResNet. Other methodologies that he applied are GoogleNet, DenseNet, VGG, AlexNet, LLC, ScSPM and accuracy achieves from these models is 86.5%, 89.1%, 87.5%, 85.6%, 64.9%, and 60.6%. By combining (HOG + SVM) and (SIFT + BOVW) Zhi received the accuracy OF

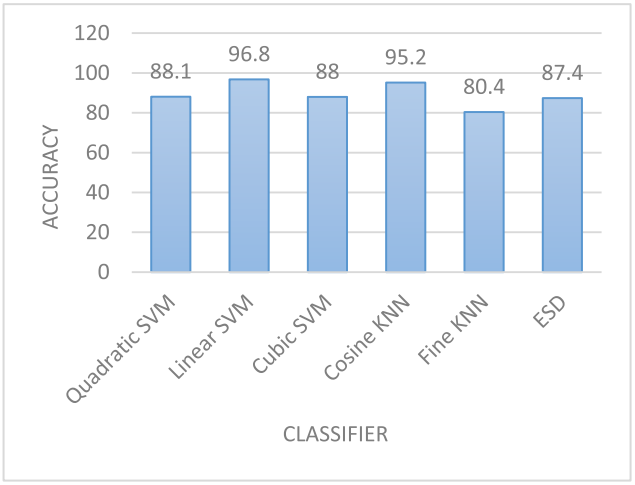


FIGURE 10. Accuracy graph of different classifiers applied on a dataset.

TABLE 5. *Zhi et al.* simulation results performed on MTARSI [14] dataset.

Algorithms	Accuracy
LLC	64.9%
SIFT +BOVW	59.0%
ScSPM	60.6%
HOG + SVM	61.3%
VGG	87.5%
GoogleNet	86.5%
AlexNet	85.6%
EfficientNet	89.7%
DenseNet	89.1%
ResNet	89.6%
Linear SVM (our proposed methodology)	**96.8%**

61.3% and 59.0%. The Overall result performed by Zhi *et al.* is shown in **Table 5**.

V. CONCLUSION

The acknowledgment of aircraft category from remotely sensed imageries draws great investigational attention. In this article, we first analyze numerous datasets of satellite images that have been widely used for object classification. We concluded that individual dataset is not either appropriate for our research area or have noteworthy limitations. Finally, we found a new benchmark dataset called MTARSI for aircraft type recognition. This dataset practically complies with the aircraft type recognition because no other dataset has this type of variety. Most of the methodologies were evaluated on diverse datasets under different investigational situations. MTARSI dataset is utilized to evaluate and review the performance of airplane type identification methodologies for natural pictures. It also gives advantages to the growth of computer vision, image manipulation, and target identification procedures for aerial imageries. After that, we performed several delegated aircraft type identification methodologies with multiple investigational conventions on the purposed novel dataset. We also notice that the data package evidently differentiates the performance of various methods. The scholars using the proposed MTARSI dataset will therefore have a sound foundation of outcomes to compare. Despite the success of the feature fusion strategy for aircraft image classification, the study was limited to satellite aircraft images. It did not investigate the influence of the multi-resolution method on aerial images. There is no information about the specific contribution of each level of resolution in the image classification task exists. In our upcoming effort, we will exploit this dataset to extend superior aircraft identification techniques. In addition, we further enlarge to accumulate more plentiful data founded on these aerial datasets and consider as the other object category of remotely sensed images.

REFERENCES

[1] K. Fu, W. Dai, Y. Zhang, Z. Wang, M. Yan, and X. Sun, "MultiCAM: Multiple class activation mapping for aircraft recognition in remote sensing images," *Remote Sens.*, vol. 11, no. 5, p. 544, Mar. 2019.
[2] J. Chen, B. Zhang, and C. Wang, "Backscattering feature analysis and recognition of civilian aircraft in TerraSAR-X images," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 4, pp. 796–800, Apr. 2015.
[3] J. Zuo, G. Xu, K. Fu, X. Sun, and H. Sun, "Aircraft type recognition based on segmentation with deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 2, pp. 282–286, Feb. 2018.
[4] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7th IEEE Int. Conf. Vis.*, Sep. 1999, pp. 1150–1157.
[5] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
[6] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2005, pp. 886–893.
[7] G. Liu, X. Sun, K. Fu, and H. Wang, "Aircraft recognition in high-resolution satellite images using coarse-to-fine shape prior," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 3, pp. 573–577, May 2013.
[8] C. Xu and H. Duan, "Artificial bee colony (ABC) optimized edge potential function (EPF) approach to target recognition for low-altitude aircraft," *Pattern Recognit. Lett.*, vol. 31, no. 13, pp. 1759–1772, Oct. 2010.
[9] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: Benchmark and state of the art," *Proc. IEEE*, vol. 105, no. 10, pp. 1865–1883, Oct. 2017.
[10] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proc. 18th SIGSPATIAL Int. Conf. Adv. Geograph. Inf. Syst.*, 2010, pp. 270–279.
[11] K. Fu, Y. Li, H. Sun, X. Yang, G. Xu, Y. Li, and X. Sun, "A ship rotation detection model in remote sensing images based on feature fusion pyramid network and deep reinforcement learning," *Remote Sens.*, vol. 10, no. 12, p. 1922, Nov. 2018.
[12] X. Yang, H. Sun, K. Fu, J. Yang, X. Sun, M. Yan, and Z. Guo, "Automatic ship detection in remote sensing images from Google earth of complex scenes based on multiscale rotation dense feature pyramid networks," *Remote Sens.*, vol. 10, no. 1, p. 132, Jan. 2018.
[13] J. G. Ha, H. Moon, J. T. Kwak, S. I. Hassan, M. Dang, O. N. Lee, and H. Y. Park, "Deep convolutional neural network for classifying Fusarium wilt of radish from unmanned aerial vehicles," *J. Appl. Remote Sens.*, vol. 11, no. 4, 2017, Art. no. 042621.

[14] Z.-Z. Wu, S.-H. Wan, X.-F. Wang, M. Tan, L. Zou, X.-L. Li, and Y. Chen, "A benchmark data set for aircraft type recognition from remote sensing images," *Appl. Soft Comput.*, vol. 89, Apr. 2020, Art. no. 106132.

[15] J. Hu, P. Ghamisi, and X. Zhu, "Feature extraction and selection of Sentinel-1 dual-pol data for global-scale local climate zone classification," *ISPRS Int. J. Geo-Inf.*, vol. 7, no. 9, p. 379, Sep. 2018.

[16] A. E. Maxwell, T. A. Warner, and F. Fang, "Implementation of machine-learning classification in remote sensing: An applied review," *Int. J. Remote Sens.*, vol. 39, no. 9, pp. 2784–2817, May 2018.

[17] B. Tu, N. Li, L. Fang, D. He, and P. Ghamisi, "Hyperspectral image classification with multi-scale feature extraction," *Remote Sens.*, vol. 11, no. 5, p. 534, Mar. 2019.

[18] S. Georganos, T. Grippa, S. Vanhuysse, M. Lennert, M. Shimoni, S. Kalogirou, and E. Wolff, "Less is more: Optimizing classification performance through feature selection in a very-high-resolution remote sensing object-based urban application," *GISci. Remote Sens.*, vol. 55, no. 2, pp. 221–242, Mar. 2018.

[19] J. Song, S. Gao, Y. Zhu, and C. Ma, "A survey of remote sensing image classification based on CNNs," *Big Earth Data*, vol. 3, no. 3, pp. 232–254, Jul. 2019.

[20] M. Afifi, "11K hands: Gender recognition and biometric identification using a large dataset of hand images," *Multimedia Tools Appl.*, vol. 78, no. 15, pp. 20835–20854, Aug. 2019.

[21] J. M. Núñez, S. Medina, G. Ávila, and J. Montejano, "High-resolution satellite imagery classification for urban form detection," in *Satellite Information Classification and Interpretation*. London, U.K.: IntechOpen, 2019.

[22] S. S. Yadav and S. M. Jadhav, "Deep convolutional neural network based medical image classification for disease diagnosis," *J. Big Data*, vol. 6, no. 1, p. 113, Dec. 2019.

[23] A. N. Parveen, H. H. Inbarani, and E. S. Kumar, "Performance analysis of unsupervised feature selection methods," in *Proc. Int. Conf. Comput., Commun. Appl.*, 2012, pp. 1–7.

[24] Y. Yu, J. Huang, S. Liu, J. Zhu, and S. Liang, "Cross target attributes and sample types quantitative analysis modeling of near-infrared spectroscopy based on instance transfer learning," *Measurement*, vol. 177, Jun. 2021, Art. no. 109340.

[25] Y. Guo, X. Jia, and D. Paull, "Effective sequential classifier training for SVM-based multitemporal remote sensing image classification," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 3036–3048, Jun. 2018.

[26] R. Li, S. Xu, Y. Zhou, S. Li, J. Yao, K. Zhou, and X. Liu, "Toward group applications of zinc-silver battery: A classification strategy based on PSO-LSSVM," *IEEE Access*, vol. 8, pp. 4745–4753, 2020.

[27] G. Alimjan, T. Sun, Y. Liang, H. Jumahun, and Y. Guan, "A new technique for remote sensing image classification based on combinatorial algorithm of SVM and KNN," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 32, no. 7, Jul. 2018, Art. no. 1859012.

[28] W. Zhou, S. Newsam, C. Li, and Z. Shao, "PatternNet: A benchmark dataset for performance evaluation of remote sensing image retrieval," *ISPRS J. Photogramm. Remote Sens.*, vol. 145, pp. 197–209, Nov. 2018.

[29] S. Maji, E. Rahtu, J. Kannala, M. Blaschko, and A. Vedaldi, "Fine-grained visual classification of aircraft," 2013, *arXiv:1306.5151*.

[30] Q. Liu, X. Xiang, Y. Wang, Z. Luo, and F. Fang, "Aircraft detection in remote sensing image based on corner clustering and deep learning," *Eng. Appl. Artif. Intell.*, vol. 87, Jan. 2020, Art. no. 103333.

[31] K. Zhao and X. Ren, "Small aircraft detection in remote sensing images based on YOLOv3," *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 533, no. 1, 2019, Art. no. 012056.

[32] B. Jiang, X. Li, L. Yin, W. Yue, and S. Wang, "Object recognition in remote sensing images using combined deep features," in *Proc. IEEE 3rd Inf. Technol., Netw., Electron. Automat. Control Conf. (ITNEC)*, Mar. 2019, pp. 606–610.

[33] Y. Li, K. Fu, H. Sun, and X. Sun, "An aircraft detection framework based on reinforcement learning and convolutional neural networks in remote sensing images," *Remote Sens.*, vol. 10, no. 2, p. 243, Feb. 2018.

[34] G. Sheng, W. Yang, T. Xu, and H. Sun, "High-resolution satellite scene classification using a sparse coding based multiple feature combination," *Int. J. Remote Sens.*, vol. 33, no. 8, pp. 2395–2412, Apr. 2012.

[35] 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.

[36] 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.

[37] L. Zhao, P. Tang, and L. Huo, "Feature significance-based multibag-of-visual-words model for remote sensing image scene classification," *J. Appl. Remote Sens.*, vol. 10, no. 3, Jul. 2016, Art. no. 035004.

[38] B. Fernando, E. Fromont, and T. Tuytelaars, "Mining mid-level features for image classification," *Int. J. Comput. Vis.*, vol. 108, no. 3, pp. 186–203, Jul. 2014.

[39] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015.

[40] K. Yu, C. Dong, C. Change Loy, and X. Tang, "Deep convolution networks for compression artifacts reduction," 2016, *arXiv:1608.02778*.

[41] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[42] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[43] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva, "Learning deep features for scene recognition using places database," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2014, pp. 487–495.

[44] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. 14th Int. Conf. Artif. Intell. Statist.*, 2011, pp. 315–323.

[45] C. Li, S. X. Yang, Y. Yang, H. Gao, J. Zhao, X. Qu, Y. Wang, D. Yao, and J. Gao, "Hyperspectral remote sensing image classification based on maximum overlap pooling convolutional neural network," *Sensors*, vol. 18, no. 10, p. 3587, Oct. 2018.

[46] M. Sharif, U. Tanvir, E. U. Munir, M. A. Khan, and M. Yasmin, "Brain tumor segmentation and classification by improved binomial thresholding and multi-features selection," *J. Ambient Intell. Humanized Comput.*, early access, pp. 1–20, Oct. 2018.

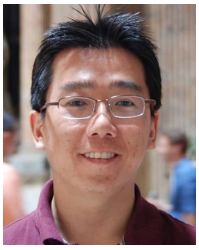
[47] T. Akram, M. A. Khan, M. Sharif, and M. Yasmin, "Skin lesion segmentation and recognition using multichannel saliency estimation and M-SVM on selected serially fused features," *J. Ambient Intell. Humanized Comput.*, early access, pp. 1–20, Sep. 2018.

FAISAL AZAM is currently working with the Department of Computer Science, COMSATS University Islamabad, Pakistan, as an Assistant Professor. His work covers agriculture image analysis, identification, classification, and prediction of plant and animal diseases using image analysis and artificial intelligence, geological analysis and prediction system using multi agent system models, and blockchain technology. His recent publications include deep learning-based technique for COVID detection from X-Ray images. His work has been published in reputed journals.

AKASH RIZVI received the M.S. degree in machine learning and image processing in 2021. He is currently a Graduate Student at COMSATS University Islamabad, Wah Campus. His research interests include features extraction (learned and low-level features), digital image processing super-resolution and reconstruction, pattern recognition, medical image analysis, deep learning for data understanding and images classifications. He is currently working on MATLAB for accomplishing research tasks. His future interest includes further research in the similar domain.

WAZIR ZADA KHAN (Senior Member, IEEE) received the bachelor's and master's degrees in computer science from COMSATS University Islamabad, Wah Campus, in 2004 and 2007, respectively, and the Ph.D. degree from the Department of Electrical and Electronic Engineering, Universiti Teknologi PETRONAS, Malaysia, in 2015. He is currently working as an Associate Professor with the Department of Computer Science, University of Wah, Wah, Pakistan. He is also serving as a Researcher at the ''Global Foundation for Cyber Studies and Research'' (https://www.gfcyber.org), which is an independent, non-profit, and non-partisan cybersecurity think-tank based in Washington D.C. He has published over 85 research papers in the journals and conferences of international repute. He has more than ten years of teaching/professional experience in Pakistan and Saudi Arabia. His current research interests include wireless sensor networks, security and privacy, blockchain, the IoT, the IIoT, and machine learning. He is serving as a reviewer for many reputed journals and also a member of the technical program committee for many international conferences.

MOHAMMED Y. AALSALEM received the master's and Ph.D. degrees from the Faculty of Engineering and Information Technology, The University of Sydney, Australia, with a specialization of sensors networks, the Internet of Things, computer networking, network security, and trust management, in March 2009. He was the Dean and the Founder of the Deanship of E-Learning and Distance Learning with Jazan University, from 2009 to 2014, where he was also the Dean of the Faculty of Computer and Information Technologies, from 2014 to August 2018. He is currently an Assistant Professor with the Faculty of Computer and Information Technologies, Jazan University.



HEEJUNG YU (Senior Member, IEEE) received the B.S. degree in radio science and engineering from Korea University, Seoul, South Korea, in 1999, and the M.S. and Ph.D. degrees in electrical engineering from the Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 2001 and 2011, respectively. From 2001 to 2012, he was with the Electronics and Telecommunications Research Institute, Daejeon; and with Yeungman University, Gyeongsan, South Korea, from 2012 to 2019. He is currently an Associate Professor with the Department of Electronics and Information Engineering, Korea University, Sejong, South Korea. His research interests include statistical signal processing and communication theory.

YOUSAF BIN ZIKRIA (Senior Member, IEEE) is currently working as an Assistant Professor with the Department of Information and Communication Engineering, Yeungnam University, South Korea. He authored more than 100 refereed articles, conference papers, book chapters, and patents. He published papers at the top venue, including IEEE COMMUNICATIONS SURVEYS & TUTORIALS, *IEEE Wireless Communications Magazine*, *IEEE Network*, *Future Generation Computer Systems* (Elsevier), and *Sustainable Cities and Society* (Elsevier). He has managed numerous FT/SI in SCI/E indexed journals. His research interests include the IoT, 5G, machine learning, wireless communications and networks, WSNs, routing protocols, CRAHN, CRASN, transport protocols, VANETs, embedded systems, and network and information security. He also held the prestigious CISA, JNCIS-SEC, JNCIS-ER, JNCIA-ER, JNCIA-EX, and Advance Routing Switching and WAN Technologies certifications. He has been listed in the world's top 2% researchers/scientists published by Stanford University and Elsevier.