

# Evaluation of GAM Classifier Performance for Airplane Remote Sensing Images Based on SIFT Features

Haidy S. Fouad <sup>1,\*</sup> and Hend A. Elsayed <sup>2</sup>

<sup>1</sup> Business Information Systems Department, Higher Institute of Advanced Administrative Sciences and Computers, Beheira, Egypt

<sup>2</sup> Electrical Engineering Department, Faculty of Engineering, Damanhour University, Egypt  
Email: haidyit@gmail.com (H.S.F.); hend.alielsayed@yahoo.com (H.A.E.)

\*Corresponding author

**Abstract**—Remote sensing technology and its applications have attracted the attention of researchers. Background variation and the small objects in remote sensing images make the classification process a challenging task. In several domains, Generalized Additive Models (GAMs) have demonstrated their ability to capture nonlinear interactions between explanatory variables and a response variable. This research evaluates the GAM with Scale Invariant Feature Transform (SIFT) for airplane remote sensing image classification. SIFT is a widely used local feature detection algorithm that performs best under scale and image rotations. We compared their performance with different methods, such as Harris-Stephens (HARRIS), Features from Accelerated Segment Test (FAST), Maximally Stable Extremal Regions (MSER), Oriented FAST Rotated BRIEF (ORB), and Binary Robust Invariant Scalable Keypoints (BRISK). To evaluate the results of the GAM with SIFT, Support Vector Machine (SVM), Discriminant Function Analysis (LDA), Quadratic Discriminant Function Analysis (QDA), and K-Nearest Neighbors (KNN) were applied. Accuracy rate, recall, precision, F-measure, and Receiver Operating Characteristic (ROC) curve were used as evaluation indexes. Based on the test dataset, SIFT features with the GAM increase precision, accuracy, recall, F-measure, and ROC curve compared to other applied classifiers. We show the performance of the applied airplane classification technique using two benchmark datasets from Google Earth, which are NWPU-RESISC-45 and UCAS-AOD.

**Keywords**—object classification, Generalized Additive Model (GAM), local feature descriptors, remote sensing images, Scale Invariant Feature Transform (SIFT)

## I. INTRODUCTION

Airplane classification is essential in various military applications, such as airport surveillance, to face challenges like security-relevant areas surrounding the monitored zone [1] and transportation activity analysis to assess airport efficiency [2]. Besides, it has many applications in the industrial and civil arenas. When

algorithms are applied to extract features from remote sensing images, there is some interference from external factors. Furthermore, images of airplanes on the ground are frequently small with multiple scales. All these factors make the classification process a challenging task. This research evaluates the Generalized Additive Model (GAM) as a classification tool based on Scale Invariant Feature Transform (SIFT) to find out an optimal method for airplane detection in remote sensing images. SIFT is the nominated feature extraction method for multiple reasons: it gains the highest score on a comparison of object recognition algorithms [3, 4], enables correct matching, is robust in extracting small objects [5], and is used in many tasks such as image registration, motion tracking, the image stitching process [6], and plant species classification [7]. Besides, we adopt the GAM since it creates a great compromise between accuracy and interpretability, which are two opposing objectives, representing a challenge in a machine learning model as it usually has to choose between accuracy and interpretability. The GAM algorithm produces more accurate results compared with other statistical algorithms such as the neural network-based method [8, 9]. Furthermore, it outperforms the mathematical calculation approach as shown in [10–12]. The GAM is widely used in a variety of fields, including business, healthcare, ecology, and climatology. In general, object detection algorithms are made up of three major modules.

### A. Proposal Generation

Proposal generation is one of the main distinctions between published object classification algorithms. It can be achieved by sliding windows, distinct points or marks, image segmentation, a Circle-Frequency Filter (CFF), or symmetric line segments [13]. Many methods have been presented to build descriptors; one method presents a log-polar grid with high central contrast, another forms them from the intensity difference between a pixel and the center pixel inside a window, and another estimates a picture's

scale and orientation first, then normalizes the image in terms of scale and rotation, and finally characterizes the

normalized image. SIFT is an example of such a representation [14].

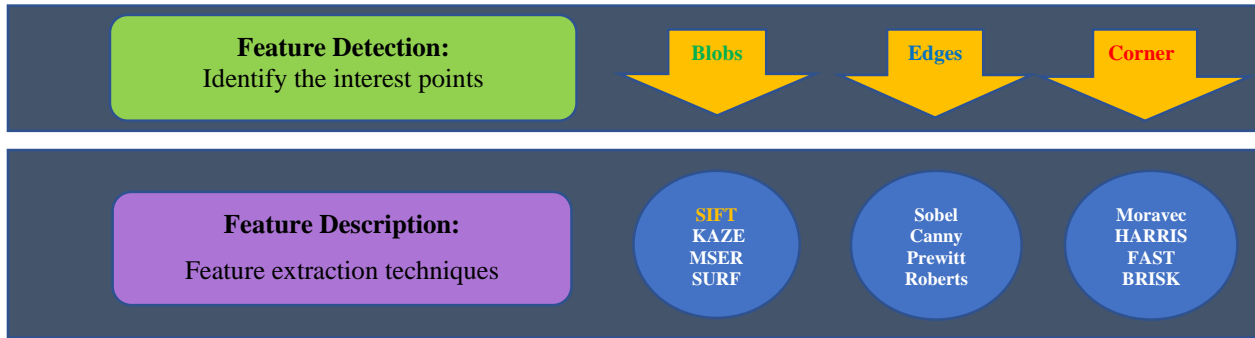


Fig. 1. A variety of detection and extraction techniques for local features.

### B. Feature Extraction

The two basic stages of obtaining local features are feature detection and feature description. Fig. 1 shows their common methodology and the related algorithms. SIFT, the method used in this research, deploys blobs for feature detection.

### C. Classification

In this research, GAM is compared with Support Vector Machine (SVM), Discriminant Function Analysis (LDA), Quadratic Discriminant Function Analysis (QDA), and K-Nearest Neighbors (KNN) models as they are the most commonly used classifiers [15, 16]. The target of GAMs is to enable the linear model to learn nonlinear relationships [17].

## II. LITERATURE REVIEW

This section is divided into two parts. The first subsection focuses on a variety of feature extraction techniques. The second subsection reviews airplane classification directions.

### A. Feature Extraction Techniques

Different methods of feature extraction have been suggested for airplane detection, for example, Yu *et al.* [18] proposed a strategy based on global and local multiscale feature fusion, Kang *et al.* [19] proposed an innovative Scattering Feature Relation Network (SFR-Net), and Ying *et al.* [20] fused local and global features. SIFT is the most accurate algorithm compared with Oriented FAST Rotated BRIEF (ORB), Speeded-up Robust Features (SURF), and Binary Robust Invariant Scalable Keypoints (BRISK) [21]. Researchers have never stopped improving SIFT algorithms such as PCA-SIFT, Global context SIFT (GSIFT), a SIFT descriptor with color invariant characteristics (CSIFT), Affine-SIFT (ASIFT), and patch-SIFT [15], but SIFT performs the best under scale, rotation, blur, and affine changes [22]. In contrast, CSIFT does not illuminate change, PCA-SIFT is always the second in different circumstances, and both GSIFT and ASIFT have lower matching correct rates.

### B. Airplane Classification Direction

Recently, there have been two directions that many have used for object classification techniques in an image: traditional and deep learning-based. Recently, many object classification algorithms have been based on deep learning frameworks [23, 24], especially Convolutional Neural Network (CNN) [25, 26]. Deploying CNN algorithms is mainly classified into two groups. The first is airplane identification based on regional proposals, such as Spatial Pyramid Pooling (SPP-Net), Region-based Fully Convolutional Networks (R-FCN), and Region-based Fully Convolutional Network (R-CNN). The second group predicts object bounding boxes for an image in a one-stage process that does not take into account the region proposal stage and runs detection directly on a dense sampling of locations, such as You only look once (YOLO) and Single Shot MultiBox Detector (SSD) [27, 28]. There are many directions for improving them [29, 30]. One-stage detectors have a high processing speed, but their accuracy is low for high-precision applications [31]. On the other hand, two-stage detectors provide high accuracy but are inefficient and more time-consuming. Another challenge is classifying objects at different scales. Although YOLO uses fast techniques, it is not perfect at being computationally intensive or sensitive to changes in lighting [32].

As shown in Table I, there is a contrast between deep learning and traditional methods. Deep neural networks require extensive training and a lot of computing power because they require many processes, such as convolutions and transformations, and on limited sample data sets, traditional machine learning has a greater solution impact. Thus, traditional computer vision techniques remain very useful even in the age of deep learning [33-34]. For airplane classification, traditional methods such as Deformable Parts Model (DPM) have been used with Histogram of Oriented Gradients (HOG), Bag-of-Words (BoW) features with a cascaded AdaBoost classifier [35], features by sparse coding and constraint pooling with SVM [36], and Gabor filter with SVM [37].

TABLE I. COMPARISON BETWEEN DEEP LEARNING AND TRADITIONAL IMAGE PROCESSING

Criteria	Traditional methods	Deep learning
Training dataset	Small	Large
Computing power	Low	High
Feature engineering	Required	Unnecessary
Training time	Short	Long
Annotation time	Short	Long
Algorithm transparency	High	Low
Domain expertise	High	Low
Priors	Few	Many
Deploying flexibility	High	Low
Expense	Low	High

### III. MATERIALS AND METHODS

#### A. Scale Invariant Feature Transform (SIFT)

SIFT first identifies what are called blobs or areas of interest at different scales. It uses a 2-D Gaussian filter smoothing over the whole photo and extracts normalized Log Gaussian values at various sigma (the value of Gaussian variance is known as sigma). It just so happens that these Gaussian values are rather helpful at differentiating blobs at various scales. It is possible to identify certain blobs at 1 sigma, others at n sigma, and so on. SIFT constructs the detectors and descriptors in four steps; Fig. 2 shows the sequence of those steps. The goal of feature detection is to identify the characteristics of the location, while the goal of feature description is to represent the feature. In this research, image features serve as words in BoW, which has a significant performance in terms of computation time and gives effective and validated results [38].

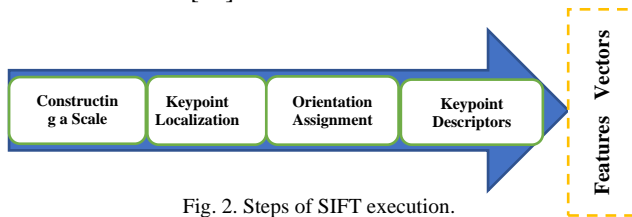


Fig. 2. Steps of SIFT execution.

##### 1) Constructing a scale space

SIFT creates blurrier versions of the original image by using the Gaussian blur technique. To ensure being scale-independent, the original image shrank to half of its original size, then produced blurry images once more, and kept doing so for multiple octaves. Scale invariance occurs when the candidate keypoint is selected from the scale with the highest frequency measurement across multiple scales [39]. After creating the scale space, the algorithm uses linear diffusion to create Difference-of-Gaussian (DoG) images for identifying blobs at various scales.

##### 2) Keypoint localization

The target is to filter the selected keypoints and get rid of unusual ones, using measures of their stability. Second, edges that are difficult to detect get rejected because they may not be noise-resistant.

##### 3) Orientation assignment

To ensure rotation invariance, orientations are assigned to each keypoint location based on local image gradient directions. In addition, an orientation histogram depicting all gradients around each keypoint is produced. The orientation for that keypoint is determined by selecting the peak in the histogram that has the dominant direction.

##### 4) Keypoint description

The target is to find out a unique fingerprint for each keypoint, so the neighborhood of every interest point is represented by a feature vector. Each point is given a 128-D feature vector by the descriptor based on 4 by 4 surrounding subregions.

#### B. Generalized Additive Model (GAM)

We use logistic GAMs in classification. The main advantage of GAMs is that they can deal with highly nonlinear and nonmonotonic relationships between the class and the predictors without the need for the explicit use of variable transformations or polynomial terms. This is because the GAM smoothing functions do these tasks automatically. In this respect, the smoothing functions are similar to the hidden layer in an artificial neural network. The flexible smoothing in GAMs is actually constructed out of many smaller functions called basis functions. Each smoothing function is the sum of a number of basis functions, and each basis function is multiplied by a coefficient [40]. For each predictor, the typical GAM employs a univariate form function, e.g., Eq. (1):

$$y \sim \text{Binomial}(n, \mu)$$

$$g(\mu) = \log \mu / (1 - \mu) = c + f_1(x_1) + f_2(x_2) + \dots + f_p(x_p) \quad (1)$$

where  $y$  is a response variable that applies the binomial distribution with the probability of positive class  $\mu$  in  $n$  observations. A logit link function is represented by  $g(\mu)$ , and  $c$  is an intercept or constant term.  $f_i(x_i)$  is the  $i$ -th predictor's univariate shape function, which is a boosted tree for a predictor tree.

#### C. Classifiers Comparison

LDA and QDA are both frequently utilized because of their ability to address various multiclass issues without the need to tweak hyperparameters to increase classification accuracy. In this research, a multivariate normal distribution is used. The classification model for LDA has the same covariance matrix for all classes; in contrast, the means and covariances change between classes for QDA. The cosine KNN is applied as it has the best accuracy among the other distance metrics, such as Euclidean and Minkowski, with 10-nearest neighbors after trying different  $k$  values. In KNN, each normalized cluster histogram serves as a point in an N-dimensional space.

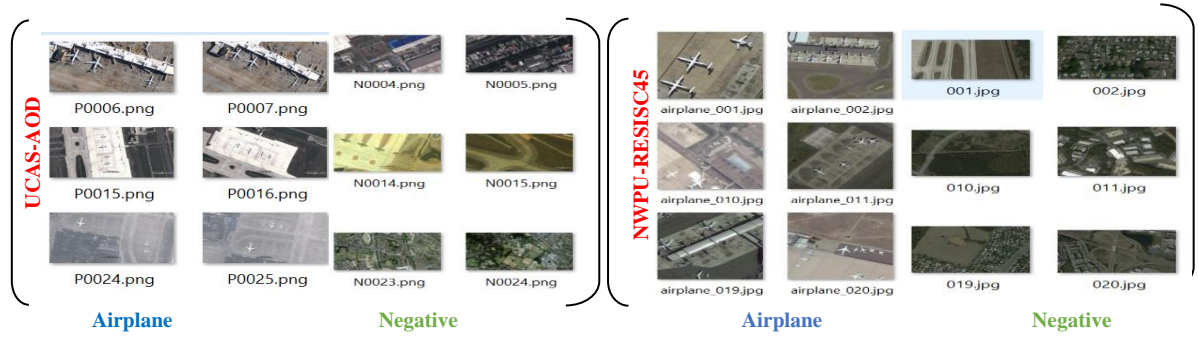


Fig. 3. Sample images of the training set in the NWPU-RESISC45 and UCAS-AOD datasets.

#### IV. DATASET DESCRIPTION AND EXPERIMENTAL SETTINGS

All the algorithms are implemented in MATLAB R2021b with Computer Vision Toolbox and Machine Learning Toolbox. They were carried out on a workstation with an Intel (R) Core (TM) i5-8265U CPU, 1.60 GHz, 1.80 GHz, and 8.0 GB of RAM.

##### A. Dataset

For this experiment, we used two benchmark datasets, which are NWPU-RESISC45 [41] and UCAS-AOD [42]. NWPU-RESISC45 is a classical remote sensing image scene classification that includes 45 classes. Each class includes 700 images. The dataset was created by Northwestern Polytechnical University (NWPU). In our experiment, we used 600 samples from airplane class and 600 samples from airport class as negative images. UCAS-AOD is a remote sensing image dataset that is on a large scale and is annotated by the University of the Chinese Academy of Sciences. It includes car and aircraft classes, as well as negative images. Fig. 3 shows samples of each dataset and illustrates the rich image variations in airplane scales and viewpoint in NWPU-RESISC45 dataset. We used 600 samples from the airplane class and 600 from the negative, with Google Earth as the data source for both of them. Table II shows the descriptions of these datasets.

TABLE II. DATASETS DESCRIPTION

Dataset	Airplane images	Total number	Object	Year
NWPU-RESISC45	700	31,500	2	2016
UCAS-AOD	1,000	2,410	45	2015

##### B. Experimental settings

In order to train and validate the proposed solutions, 500 images are used. The positive image set contains 250 images that have at least one airplane in each, and the 250

images that represent airport images do not contain any targets. For the testing process, 100 new images have been used.

In this research, the study’s analysis steps are applied with MATLAB, as shown in Fig. 4, and can be listed as follows:

Input: Manage 500 images of our datasets as an object for training and validation, and 100 images for testing as shown in Fig 5, image preprocessing has been done by converting the images to grayscale and resizing all the images to 256×256.

*Step 1:* Extract the features’ descriptors by creating the Bag-of-SIFT model. The bag has an extractor function, which contains detect and extract functions.

*Step 2:* Encode and compute vocabulary for each image with BoW into a feature vector by computing the histogram of visual word occurrences for each image.

*Step 3:* Train and validate the dataset with the model classifier using GAM, KNN, SVM, QDA, and LDA.

*Step 4:* Test the performance of the model. Classify the testing dataset using an ensemble of classification models.

*Step 5:* Evaluate the results using accuracy, precision, recall, and the F1-Score. The formulas for these metrics are listed in Eqs. (2)–(5):

$$Accuracy = TP + \frac{TN}{TP} + TN + FP + FN \quad (2)$$

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

$$F1-Score = \frac{2 \times (precision \times recall)}{precision + recall} \quad (5)$$

*Step 6:* Test the accuracy of the models that were trained on the entire dataset, including training and validation data, using predict functions.

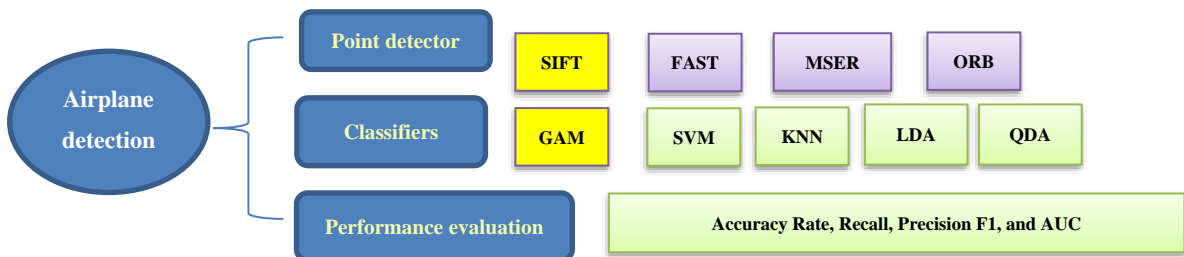


Fig. 4. Airplane classification methodology.



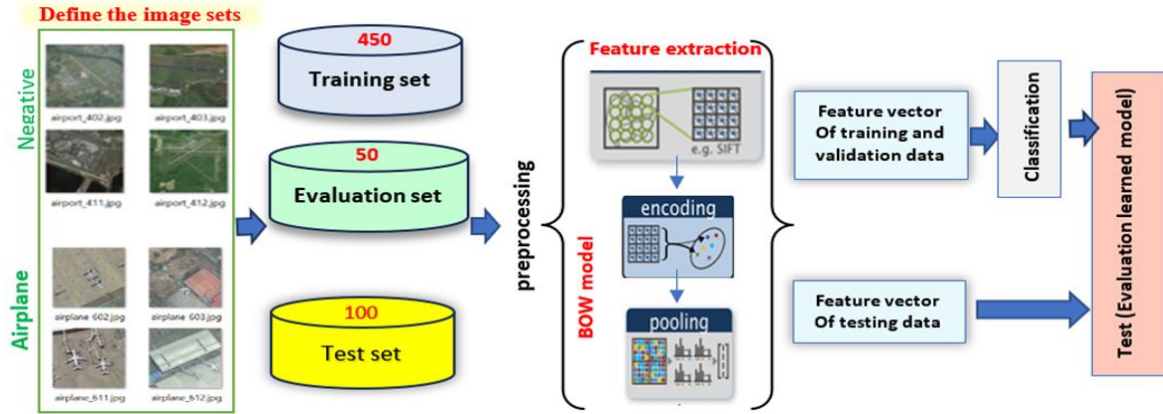


Fig. 5. Illustration of the proposed method process.

### C. Experiment Parameter Settings

The goal of the parameter modification is to improve GAM’s accuracy based on the dependencies between the parameters and accuracy. The accuracy is good as the data contain the same numbers of positive and negative images. The accuracy of all methods was compared over multiple runs, and the best was chosen.

#### 1) Classifier parameters

GAM, SVM, KNN, QDA, and LDA were used in classification based on SIFT descriptors. After trying a different number of fold settings, 10-fold cross-validation was used as it increases the accuracy. With SVM, the optimal kernel function was the Radial Basis Function (RBF). In KNN, after comparing different distinction models between classes such as cubic, cosine, and weighted, the cosine distance metric got the best accuracy with 10-nearest neighbors after trying different values of nearest neighbors. Finally, QDA and LDA used the full covariance structure.

#### 2) Bag of visual words parameter

The branching factor and the fraction of the strongest features are two effective parameters. The best SIFT-GAM accuracy of 86.4 was achieved with 100 branching factors and the strongest feature percentage of 80%, as shown in Figs. 6 and 7, respectively.

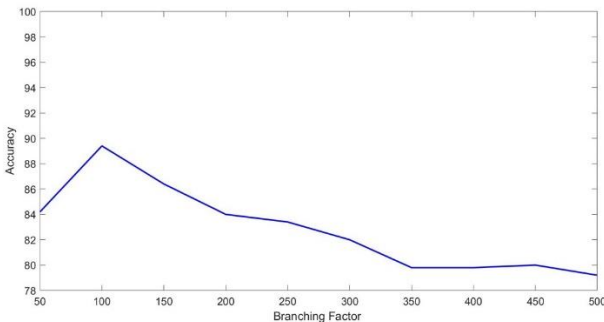


Fig. 6. The effect of the branching factor on GAM’s accuracy.

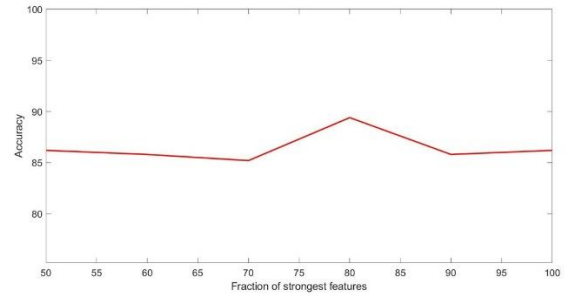


Fig. 7. The effect of the strongest feature percentage on GAM’s accuracy.

#### 3) SIFT parameters

To define the optimal values, we use empirical analysis to see the effects of changing SIFT parameters on GAM’s accuracy. Parameters like sigma, the number of scales sampled in each octave, and the contrast threshold have a powerful effect [43]. The optimal SIFT parameter values resulting from GAM are listed in Table III.

TABLE II. PARAMETERS OF SIFT

Parameter	Starting value	Step size	Final value	Optimal	Default
Contrast threshold	0	0.01	0.08	<b>0.03</b>	0.0133
Sigma	1	0.1	2	<b>1.3</b>	1.6
Strongest point	50	50	2000	<b>450</b>	All
Edge threshold	5	5	20	<b>10</b>	10

### D. Result and Discussion



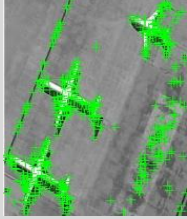

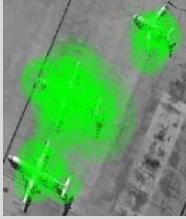
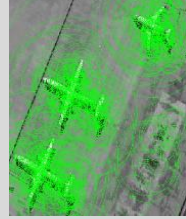
The results found can be divided into two orientations: results extracted from descriptor technique comparisons applied with GAM and results from the comparison between classifiers using SIFT.

As the NWPU-RESISC45 dataset classification results are better than those of UCAS-AOD, as shown in Table IV, the research will focus on this dataset. NWPU-RESISC45 has a variety of scales and directions for airplanes, so the discussion will concentrate on the results of NWPU-RESISC45.

TABLE IV. CLASSIFICATION ACCURACY OF THE DATASETS USING DIFFERENT CLASSIFIERS IN PERCENTAGE

Classifier	GAM	Cosine KNN	Quadratic SVM	LDA	QDA
UCAS-AOD	81.4	77.4	82	78	74
NWPU-RESISC45	89.4	86.2	87.4	88.8	83.2

TABLE V. COMPARISON OF THE APPLIED FEATURE EXTRACTION METHODS

Blobs	Corners-Single-scale detection		Regions of uniform intensity	Corners-Multiscale detection	
SIFT	HARRIS	FAST	MSER	ORB	BRISK
					
89.20%	85.20%	80.60%	81.00%	82.60%	73.60%

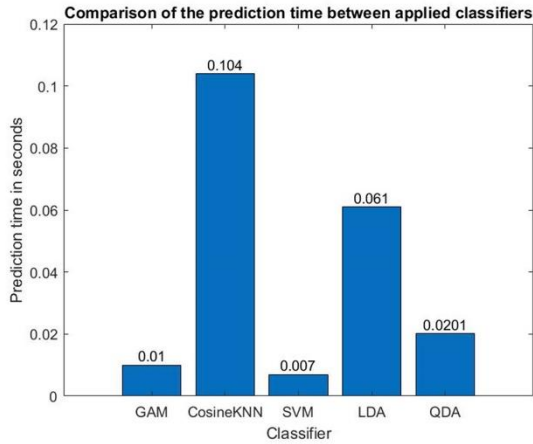


Fig. 8. Comparison of the prediction time in seconds between the applied classifiers.

### F. Classification Results Evaluation

The accuracy, recall, precision, and F1-Score are the most common measurements to evaluate the performance of classification [27]. These parameters are calculated using the confusion matrix, which contains the numbers of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) values. Table IV shows the confusion matrix results obtained using the applied classifiers. Accuracy focuses on the overall correctly classified observations, while precision is a measure of result relevancy, and recall is a measure of how many truly relevant results are returned. The F1-Score is calculated as the harmonic mean of precision and recall, which evaluates a model’s predictive abilities by analyzing its performance

### E. Descriptor Technique Results

For comparison purposes, we applied HARRIS, FAST, MSER, ORB, and BRISK in addition to SIFT. Table V shows the results in terms of accuracy and visualizes 450 of the strongest points and types of features.

in each class separately, rather than considering total performance as accuracy does. Based on SIFT features classification, GAM has the best results in accuracy, recall, and F1-Score with 500 samples. Fig. 8 shows the comparison between them in terms of accuracy, recall, precision, and F1-Score in percentage.

After we assess the models based on their validation accuracy, we test the models with 100 new images. Fig. 9 shows the time that was consumed to predict the new data. The best classification time is obtained by SVM, which requires 0.007 s. To evaluate the classification accuracy, six basic measures are used: accuracy, recall, precision, F1-Score, Area under the Curve (AUC), and ROC (receiver operating characteristic). The results shown in Fig. 9 demonstrate that GAM has the best results in accuracy, recall, and F1-Score with 500 samples. Table VI shows the number of TP, TN, FP, and FN values for each classifier. To summarize GAM’s performance, Fig. 10 shows the confusion matrix chart of SIFT feature classification with GAM, which compares the predicted labels of a classification model to the real labels. Furthermore, Fig. 11 shows the comparison between the applied classifiers using 100 new test samples.

TABLE VI. CORRECTLY AND INCORRECTLY CLASSIFIED VALIDATION DATA INSTANCES

Classifiers	TP	TN	FP	FN
GAM	219	224	31	26
KNN	227	204	23	46
SVM	223	214	27	36
LDA	223	221	27	29
QDA	200	216	50	34

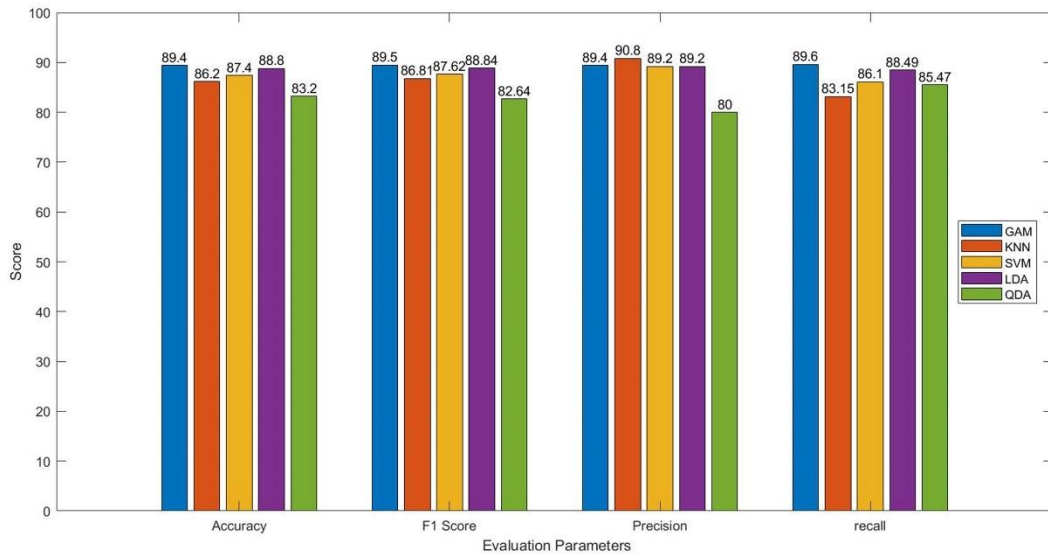


Fig. 9. Classifiers performance on the validation data in terms of accuracy, precision, recall, and F1-Score.



Fig. 10. Confusion matrix chart of SIFT feature classification with GAM.

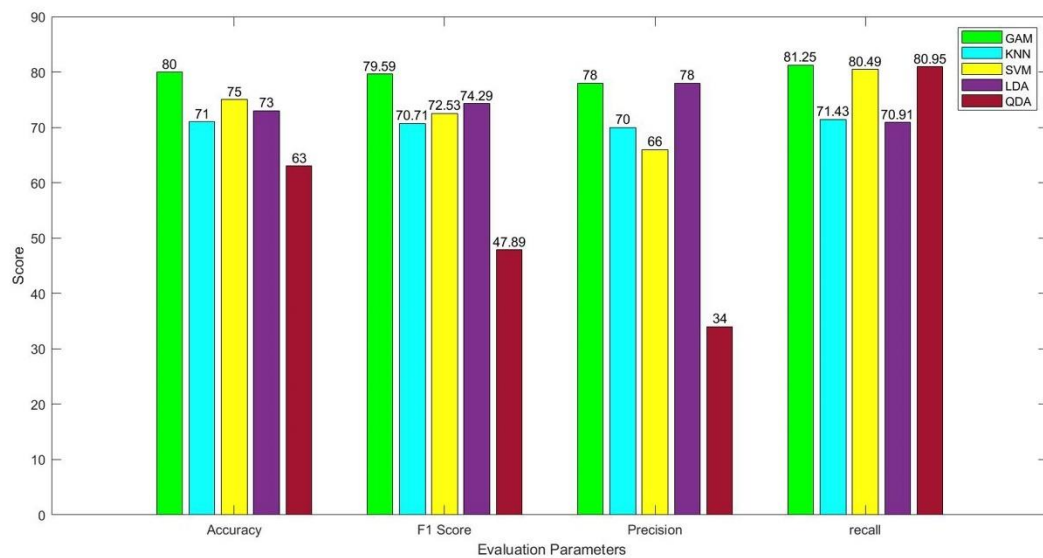


Fig. 11. Classifiers performance on the test set in terms of accuracy, precision, recall, and F1-Score.

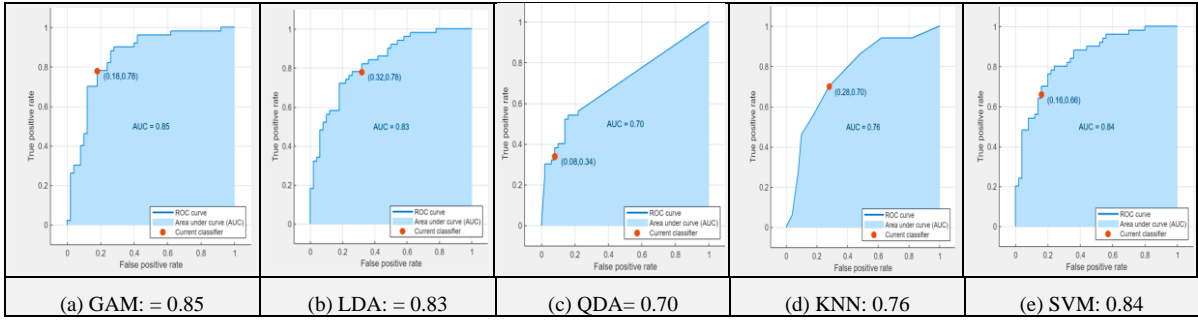


Fig. 12. Comparison between the applied classifiers using test data in terms of AUC.

GAM has the best-performing models based on their test dataset. Table VII shows the correctly and incorrectly classified test instances.

TABLE VII. CORRECTLY AND INCORRECTLY CLASSIFIED TEST INSTANCES

Classifiers	TP	TN	FP	FN
GAM	39	41	11	9
KNN	35	36	15	14
SVM	33	42	17	8
LDA	39	34	11	16
QDA	17	46	33	4

AUC and ROC curves are two machine learning metrics that are used to assess how well binary classification models perform. ROC is a plot of the proportion of True Positives (TPR) versus the proportion of False Positives (FPR) at different probability cutoffs. TPR is also called recall or sensitivity. The formulas for FPR are listed in Eqs. (6)–(8):

$$FPR = 1 - Specificity \quad (6)$$

$$FPR = \frac{FP}{TN} + FP \quad (7)$$

$$Specificity = \frac{TN}{TN+FP} \quad (8)$$

AUC represents the probability that a randomly selected positive sample will be ranked higher by the model than a randomly selected negative. The higher the AUC score, the better the model. In general, an AUC of 0.8 to 0.9 is considered excellent. GAM has the biggest AUC value, as shown in Fig. 12, which means it has the best capability for differentiating between classes with an AUC of 0.85. It indicates that there is an 85% probability that the model will be able to discriminate between classes that are positive and negative. This is due to the classifier detecting more TP and TN than FN and FP values.

The results of GAM’s performance may be affected by using the BoW technique. The histogram is the base to train a classifier as well as to classify images. Fig. 13 shows a histogram of visual word occurrences for one training image. For improving the results, future work can extend to using another image representation method.

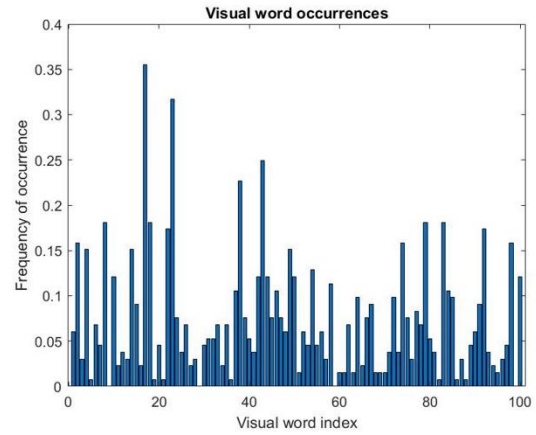


Fig. 13. The histogram of visual word occurrences.

GAM’s results capture nonlinear interactions between explanatory variables and a response variable. This can represent weaknesses when it comes to effectively capturing nonlinear patterns which require larger sample sizes. Thus, we will study the effect of data size on GAMs in the future work.

## V. CONCLUSION

In this paper, we proposed a technique for locating airplanes based on GAM using a bag of SIFT with a 10-fold stratified cross-validation technique. The first objective was to compare the results of SIFT based on the GAM classifier with different local feature detection methods, such as HARRIS, FAST, MSER, ORB, and BRISK. GAM achieved the highest accuracy, with SIFT reaching 89.4%. The second goal was to assess the effectiveness of the suggested strategy. We compared GAM-SIFT results with KNN, SVM, LDA, and QDA using accuracy, recall, precision, F1-Score, and AUC. According to the evaluation, GAM was better than the others in terms of accuracy, F1-Score, and recall with values of 89.4, 89.5, and 89.6, respectively, using the validation set, and the best of accuracy, recall, precision, F1-Score, and AUC using the test set with values of 80, 79.59, 78.00, 81.25, and 0.85, respectively. With the test set, GAM had an acceptable prediction time of 0.010 s. Experimental results demonstrated that the GAM classifier had a good performance, but with a performance close to that of other classifiers used in this research.



In this paper, we focus on the traditional learning-based methods to classify airplane satellite images. Deep learning represents another working direction as it requires huge data, extensive training, and a lot of computing power because it requires many processes, such as convolutions and transformations. In future work, we will deploy deep learning-based models by designing different Deep Convolutional Neural Networks (DCNN) for multiple objects using several layers, multiple filters, and parameters and compare their performance with other deep learning models. We will also study the effect of data size on GAM's results. Future work can extend to using other feature extraction techniques and image representation methods as well.

#### CONFLICT OF INTEREST

The authors declare no conflict of interest.

#### AUTHOR CONTRIBUTIONS

Hend A. Elsayed presented the idea and reviewed the article. The research article was written by Haidy S. Fouad, who also examined and analyzed the results. All authors approved the final version of the paper.

#### REFERENCES

- [1] L. Sharma and P. Garg, *From Visual Surveillance to Internet of Things: Technology and Applications*, CRC Press, UK, 2019.
- [2] E. Mangortey, T. G. Puranik, O. J. Pinon-Fischer *et al.*, "Classification, analysis, and prediction of the daily operations of airports using machine learning," in *Proc. AIAA Scitech 2020 Forum*, 2020, 1196.
- [3] O. Andersson and S. A. R. Marquez, "A comparison of object detection algorithms using unmanipulated testing images: Comparing SIFT, KAZE, AKAZE and ORB," Degree Project, Dept. of Computer Science, KTH Royal Institute of Technology, Stockholm, 2016.
- [4] E. Karami, S. Prasad, and M. Shehata, "Image matching using SIFT, SURF, BRIEF and ORB: Performance comparison for distorted images," arXiv preprint, arXiv:1710.02726, 2017.
- [5] Y. Yang, Y. Zhuang, H. BiShi *et al.*, "M-FCN: Effective fully convolutional network-based airplane detection framework," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 8, pp. 1293–1297, 2017.
- [6] J. Zhang and Y. Xiu, "Image stitching based on human visual system and SIFT algorithm," *The Visual Computer*, pp. 1–13, 2023.
- [7] J. Thomkaew and S. Intakosum, "Plant species classification using leaf edge," *Journal of Image and Graphics*, vol. 11, no. 1, 2023.
- [8] H. Solanki, D. Bhatpuria, and P. Chauhan, "Applications of Generalized Additive Model (GAM) to satellite-derived variables and fishery data for prediction of fishery resources distributions in the Arabian Sea," *Geocarto International*, vol. 32, no. 1, pp. 30–43, 2017.
- [9] T. Ouarda, C. Charron, P. Marpu *et al.*, "The generalized additive model for the assessment of the direct, diffuse, and global solar irradiances using SEVIRI images, with application to the UAE," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 4, pp. 1553–1566, 2016.
- [10] N. Chamidah, K. H. Gusti *et al.*, "Improving of classification accuracy of cyst and tumor using local polynomial estimator," *Telkonnika, Telecommunication Computing Electronics and Control*, vol. 16, no. 3, pp. 1492–1500, 2019.
- [11] G. S. Marcillo, N. F. Martin Diers, B. W. Diers *et al.*, "Implementation of a Generalized Additive Model (GAM) for soybean maturity prediction in African environments," *Agronomy*, vol. 11, no. 6, 1043, 2021.
- [12] A. S. Masrani, N. R. Nik Husain, N. R. Musa *et al.*, "Prediction of dengue incidence in the northeast Malaysia based on weather data using the generalized additive model," *BioMed Research International*, 2021.
- [13] U. Alganci, M. Soydas, and E. Sertel, "Comparative research on deep learning approaches for airplane detection from very high-resolution satellite images," *Remote Sens.*, vol. 12, no. 3, 2020.
- [14] A. A. Goshtasby, "Image descriptors," in *Proc. Image Registration Principles, Tools and Methods*, 2012, Springer London, ch. 5, pp. 219–246. [https://doi.org/10.1007/978-1-4471-2458-0\\_5](https://doi.org/10.1007/978-1-4471-2458-0_5)
- [15] K. Raju, B. C. Rao, K. Saikumar *et al.*, "An optimal hybrid solution to local and global facial recognition through machine learning," in *A Fusion of Artificial Intelligence and Internet of Things for Emerging Cyber Systems*, Springer, 2022, pp. 203–226.
- [16] M. Hema, "Object Detection and learning via feature descriptors," PhD thesis. Dept. of computer applications, Madurai Kamaraj Univ., Madurai, 2021.
- [17] Y. Lou, R. Caruanaand, and J. Gehrke, "Intelligible models for classification and regression," in *Proc. the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Beijing, China, 2012, pp. 150–158.
- [18] L. Yu, H. Hu, Z. Zhong *et al.*, "GLF-Net: A target detection method based on global and local multiscale feature fusion of remote sensing aircraft images," *IEEE Geoscience and Remote Sensing Letters*, vol. ED-19, pp. 1–5, 2022.
- [19] Y. Kang, Z. Wang, J. Fu *et al.*, "SFR-Net: Scattering feature relation network for aircraft detection in complex SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. ED-60, pp. 1–17, 2021.
- [20] J. Ying, H. Li, H. Yang *et al.*, "Small aircraft detection based on feature enhancement and context information," *Journal of Aerospace Information Systems*, vol. 20, no. 3, pp. 140–151, 2023.
- [21] A. J. Oliveira, B. M. Ferreira, and N. A. Cruz, "A performance analysis of feature extraction algorithms for acoustic image-based underwater navigation," *Journal of Marine Science and Engineering*, vol. 9, no. 4, 361, 2021.
- [22] J. Wu, Z. Cui, V. Sheng *et al.*, "A comparative study of SIFT and its variants," *Measurement Science Review*, vol. 13, no. 3, 2013.
- [23] J. Ren and Y. Wang, "Overview of object detection algorithms using convolutional neural networks," *Journal of Computer and Communications*, vol. 10, no. 1, pp. 115–132, 2022.
- [24] I. Iqbal, M. Younus, K. Walayat, M. U. Kakar, and J. Ma, "Automated multi-class classification of skin lesions through deep convolutional neural network with dermoscopic images," *Computerized Medical Imaging and Graphics*, vol. 88, 101843, 2021.
- [25] I. Iqbal, K. Walayat, M. U. Kakar, and J. Ma, "Automated identification of human gastrointestinal tract abnormalities based on deep convolutional neural network with endoscopic images," *Intelligent Systems with Applications*, vol. 16, 200149, 2022.
- [26] B. Azam, M. J. Khan, F. A. Bhatti *et al.*, "Aircraft detection in satellite imagery using deep learning-based object detectors," *Microprocessors and Microsystems*, vol. ED-94, 104630, 2022.
- [27] S. Zhang, X. Han, Y. Zhang *et al.*, "Aircraft detection from large-scale remote sensing images based on visual saliency and CNNs," *International Journal of Remote Sensing*, vol. 43, no. 5, pp. 1749–1769, 2022.
- [28] A. Attarian, M. Luo, Y. Luo *et al.*, "Airplane detection and classification based on mask R-CNN and YOLO with feature engineering," in *Proc. SAI Intelligent Systems Conference*, Springer International Publishing, 2022, pp. 752–768.
- [29] R. Luo, L. Chen, J. Xing *et al.*, "A fast aircraft detection method for SAR images based on efficient bidirectional path aggregated attention network," *Remote Sensing*, vol. 13, no. 15, 2940, 2021.
- [30] J. Xing, R. Luo, L. Chen *et al.*, "A comparison of deep neural network architectures in aircraft detection from SAR imagery," in *Proc. Synthetic Aperture Radar (SAR) Data Applications Cham*, Springer International Publishing, 2023, pp. 91–111.
- [31] J. Kaurand and W. Singh, "Tools, techniques, datasets and application areas for object detection in an image: A review," *Multimedia Tools and Applications*, vol. 81, no. 27, pp. 38297–38351, 2022.
- [32] L. Ge, D. Dan, and H. Li, "An accurate and robust monitoring method of full-bridge traffic load distribution based on YOLO-v3 machine vision," *Structural Control and Health Monitoring*, vol. 27, no. 12, 2020.
- [33] P. Wang, E. Fan *et al.*, "Comparative analysis of image classification algorithms based on traditional machine learning and

- deep learning,” *Pattern Recognition Letters*, vol. 141, pp. 61–67, 2021.
- [34] N. O’Mahony, S. Campbell, A. Carvalho *et al.*, “Deep learning vs. traditional computer vision,” in *Proc. Advances in Computer Vision: 2019 Computer Vision Conference (CVC)*, Springer International Publishing, 2020, vol. 1, no. 1, pp. 128–144.
- [35] F. Bi, Z. Yang, M. Lei *et al.*, “Airport aircraft detection based on local context DPM in remote sensing images,” in *Proc. IGARSS, IEEE International Geoscience and Remote Sensing Symposium*, Japan: IEEE, 2019, pp. 1362–1365.
- [36] H. Mu, H. Ni, M. Zhang *et al.*, “Tree leaf feature extraction and recognition based on geometric features and HAAR wavelet theory,” *Engineering in Agriculture, Environment and Food*, vol. 12, no. 4, pp. 477–483, 2019.
- [37] L. Cui, F. Meng, Y. M. Li Shi *et al.*, “A hierarchy method based on LDA and SVM for news classification,” in *Proc. 2014 IEEE International Conference on Data Mining Workshop*, China: IEEE, 2014, pp. 60–64.
- [38] W. A. Qader, M. M. Ameen, and B. I. Ahmed, “An overview of bag of words: Importance, implementation, applications, and challenges,” in *Proc. 2019 International Engineering Conference (IEC)*, Iraq, IEEE, 2019, pp. 200–204.
- [39] J. Almeida, R. Torres, and S. Goldenstein. (2009). SIFT applied to CBIR. *Revista de Sistemas de Informacao da FSMA*. [Online]. 4. pp. 41–48. Available: [https://www.fsma.edu.br/si/edicao4/FSMA\\_SI\\_2009\\_2\\_Principal\\_2.pdf](https://www.fsma.edu.br/si/edicao4/FSMA_SI_2009_2_Principal_2.pdf)
- [40] S. N. Wood, *Generalized Additive Models: An Introduction with R*, CRC press, UK, 2017.
- [41] G. Cheng, J. Han, and X. Lu, “Remote sensing image scene classification: Benchmark and state of the art,” *Proceedings of the IEEE*, vol. 105, no. 10, pp. 1865–1883, 2017.
- [42] H. Zhu, X. Chen, W. Dai *et al.*, “Orientation robust object detection in aerial images using deep convolutional neural network,” in *Proc. 2015 IEEE International Conference on Image Processing (ICIP)*, IEEE, 2015, pp. 3735–3739.
- [43] A. Sima and S. Buckley, “Optimizing SIFT for matching of short wave infrared and visible wavelength images,” *Remote Sens. (Basel)*, vol. 5, no. 5, pp. 2037–2056, 2013.

Copyright © 2024 by the authors. This is an open access article distributed under the Creative Commons Attribution License ([CC BY-NC-ND 4.0](https://creativecommons.org/licenses/by-nc-nd/4.0/)), which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.