Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 43 Issue: 08-2024
DOI: 10.5281/zenodo.13292340

BIRDS RECOGNITION USING DEEP LEARNING BASED ResNet-50 MODEL WITH SVM CLASSIFIER

SAIMA ANWAR LASHARI
College of Computing and Informatics, Saudi Electronic University, Riyadh, KSA.
Email: s.lashari@seu.edu.sa

Abstract

Due to anomalous changes and for the management of the environment, monitoring and surveillance of animal species is a crucial duty. Because increasing environmental change puts many animal species in danger, a detailed understanding of the intricacy of the natural environment would be preferable. Installing technological systems with the knowledge to address the issues mentioned earlier is necessary for the best protection of wild animals and the best monitoring of livestock. This article offers the ResNet50 model, built on deep learning and trained using SVMs as a classifier. This work combines the SVM classifier with several deep-learning models, including VGG_16, VGG_19, ResnNet101, DensNet121, and Mobil Net. Similar to the previous approach, this one uses support vector machines to classify birds after extracting their characteristics. The SoftMax layer is swapped out for SVM in the suggested techniques. Using images of birds, the performance of the employed models is evaluated. Support vector machines are utilized to categorize the appearances of birds in images after a ResNet50 architecture is developed to simulate their looks. Instead of using the SoftMax approach, this technique supports vector machines for a more reliable and accurate categorization of bird images. Then, another deep learning approach was created based on support vector machines, including VGG_16, VGG_19, ResnNet101, DensNet121, and Mobil Net. The suggested approaches offer improved accuracy compared to conventional methods because support vector machines are more potent classifiers than SoftMax. The performance results demonstrate that the suggested models perform better than the standard deep learning, such as VGG_16, VGG_19, and ResNet101, DensNet121, and Mobil Net-based SVM models.

Keywords: Wildlife, SVM, Resnet50, VGG_16, VGG_19, DensNet121, SoftMax, ResnNet101, DensNet121 and Mobil Net.

1. INTRODUCTION

Monitoring and surveilling wild animals in national parks is crucial for biology and ecology. Due to unusual changes and for the management of the ecosystem, it is crucial to regularly monitor wildlife species. Due to environmental change and fast habitat loss, many wildlife species are in danger [1], it will be better to have detailed information about the complexity of the natural ecosystem, the number, location, and behavior of the animals.

Identifying their habitat populations and understanding the complex natural ecosystem is important for protecting and managing ecosystems because these ecosystems can significantly affect human health [1, 2]. For proper protection of wild animals and the best monitoring of livestock, there is a need for technological systems to be installed with the expertise to combat the above-stated problems.

Different researchers create different systems, but the technological solutions that are employed to address these issues include computer vision techniques combined with neural network systems [2]. As a result, many counting techniques have been employed

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 43 Issue: 08-2024
DOI: 10.5281/zenodo.13292340

to estimate animal numbers. However, the most alluring strategy for drawing researchers' attention in the literature on automated bird recognition using aerial pictures has mostly employed representational image-processing techniques. Several spectral thresholding and filtering techniques are used by Gilmer et al. [3], Cunningham et al. [4], and Trathan [5].

Abd-Elrahman [6] developed a template-matching method for bird detection. Where else Liu et al. [7] used filtering methods and unsupervised classification for bird counting. However, despite their success in counting birds, these techniques have several drawbacks, such as the studies' use of a very small number of photos. Additionally, most are only employed to capture a small number of photos of particular species in particular environments. Due to these restrictions, it is difficult to detect birds that are dispersed throughout different settings [8].

Due to the aforementioned problems, deep learning-based techniques are increasingly being studied for use in various object detection applications. Additionally, because to their use of deep-layer learning, deep learning-based algorithms exhibit good performance on enormous amounts of data. In addition, feature selection is a full task that is needed for distinct objection detection in machine learning techniques. Deep learning-based techniques, however, can independently extract characteristics from data [9].

Convolutional neural networks (CNN) are the most widely used model among deep-learning-based object-detection techniques, having been built primarily as classification networks suitable for image-type data [10]. However, various object detection methods are developed by various research which is based on CNN based, such as Region-based Convolutional Neural Network (R-CNN) developed by [11], Fast R-CNN suggested by [12], and Faster R-CNN proposed by [13].

All these object-detection CNN-based approaches comprise two steps: the first is to identify the bounding box, and the second is a classification performed sequentially. Various methods are used for one-stage object detection, such as You Only Look Once (YOLO) [14], Single Shot MultiBox Detector (SSD) [15], and Retina Net [16]. All these methods of parallel processes were bounding boxes and classification. The one-stage object detection processing techniques are quicker than the two-stage techniques, but the accuracy performance and computing speed of these two techniques are very different from one another because the performance of these models depends on the kind of CNN architecture such as, Google net (Inception) [17], Squeeze net [18], Alex net [19], VGGNet [20], ResNet [21], or Dense net [22].

All of these object recognition and classification techniques based on deep learning outperform machine learning-based models. A number of changes have been made to further enhance the effectiveness of these techniques. Therefore, most research used deep learning-based models in various applications. Ammour *et al.* [9] used CNN based support vector machine (SVM) model for car detection using aerial photographs. Chang *et al.* [23] developed the YOLO v.2 model from pedestrian detection from aerial photographs, and Chen *et al.* [24] evaluate the effectiveness of the Faster R-CNN

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 43 Issue: 08-2024
DOI: 10.5281/zenodo.13292340

approach for identifying airports from aerial photos. Similar to this, various studies are being carried out to examine the usage of deep learning applications for wildlife monitoring with the aid of aerial pictures. In order to evaluate the effectiveness of the employed model, Maire *et al.* [25] employed CNN and simple linear iterative clustering (SLIC) techniques, and Guirado *et al.* [26] suggested a CNN-based technique for counting and detecting the presence of whales using satellite photos.

This work uses SVM models for deep learning to recognize various birds in photos. Five distinct deep-learning-based object-detection techniques were used to assess the aerial pictures of wild birds and bird decoys in diverse South Korean locations, such as lakes, beaches, reservoirs, and farms. The suggested bird detection models' accuracy, precision, and recall were also confirmed. Three key contributions to the field of visual recognition of birds are made in this paper.

1. ResNet-50 based SVM model is proposed for automatic birds' recognition.
2. Different variants of deep learning-based SVM models are compare.
3. The effectiveness of the suggested strategies is assessed in terms of accuracy, loss, precision, recall, and F1-Score against a variety of other models, including Mobile Net-based SVM models, ResNet101, DensNet121, VGG-16, and VGG-19.

The rest of the paper is organized as given. Section 2 will give the Methods and Material. The subsection 2.3 explains the proposed model in this paper. Further, section 3 elaborated on the result and discussion of the paper. Finally, the last section gives the conclusion of this research.

2. METHODS AND MATERIAL

2.1 Residual Network (ResNet-50)

The ResNet network model was developed by Kaiming [21, 27]. This model consists of designing ultra-deep networks that solve the vanishing gradient problem that exists in the previous model. Various researchers have developed ResNet with many different numbers of layers, such as 101, 34, 50, 152, and even 1202. One of the most efficient models among all of these models is the ResNet-50 network model. There are 50 layers in the ResNet-50 model. One fully linked layer is employed at the network's end, while 49 of the 50 levels are convolution layers. There were 3.9 million MACs and 25.5 million weights in the entire network.

Figure 1 depicts the fundamental block diagram of the residual block in the ResNet architecture. A residual connection makes up each of the fundamental residual blocks of the ResNet model. After executing a variety of processes (for example, convolution with various filter sizes, Batch Normalization (BN) followed by an activation function, such as a ReLU), the output of a residual layer, which originates from the output of the preceding layer can be specified depending on the outputs. However, the operations in the residual block can vary depending on residual network architectures [21].

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 43 Issue: 08-2024
DOI: 10.5281/zenodo.13292340

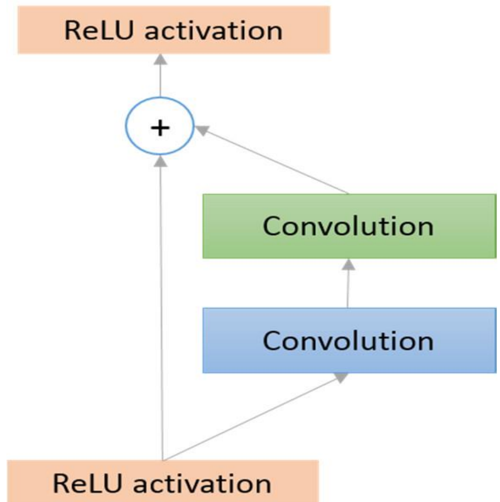

Figure 1: Residual block Diagram

2.2 Support Vector Machine (SVM)

The support vector machine technique is based on statistical theory and is used to widely solve the problem of classification [28]. At first, SVM was used for binary classification, but currently, it is used to deal with more than two classes. SVM is used to find a linear optimal hyperplane that distributes the data into two or more classes. All those elements that belong to the same class are on the same side [29].

This paper shows the impact of the SVM parameter on a bird's classification performance. The proposed technique improves the bird's classification performance of SVM to find the optimal parameters. SVM achieves data discrimination by applying kernel functions to transfer the input space to a high-dimensional feature space. It takes less time to train an SVM with a linear kernel than another kernel.

The SVM classifier's brief description is provided below. The inputs for the SVM's training set include $T = \{(x_{i,}y_i) \ldots \ldots . . (x_{l,}y_l)\}$, where $x_i \in R^2$ and $y_i \in \{-1, 1\}$. The main goal is to correctly divide the training dataset into two categories while optimizing the SVM, finding a dividing hyperplane, and doing so. The SVM optimization problem is constructed as follows:

$$\min_{w,\ b,\ \xi} \frac{1}{2}\|x_i\|^2 + c\sum_{i=1}^{l}\xi_i \tag{1}$$

$$\text{s.t} \qquad y_{i\left((w\,.\,\phi(x_i))+b\right)\geq 1-\xi_i} \qquad \xi_i \geq 0,\ \ i = 1 \ldots \ldots \ldots \ldots \ldots . . l$$

C > 0 is the error term's penalty parameter; (). The normal vector and the offset of the separating hyperplane, respectively, are denoted by w and b in Equation (1);

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 43 Issue: 08-2024
DOI: 10.5281/zenodo.13292340

Equation (1) is converted to the Lagrange dual problem as follows.

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} y_i y_j \alpha_i \alpha_j \left(\varphi(x_i).\varphi(x_j)\right) - \sum_{j=1}^{l} \alpha_j$$

$$\text{s.t} \qquad \sum_{i=1}^{l} y_i \alpha_i = 0 \qquad 0 \le \alpha_i \le C, \qquad i = 1, \dots \dots . . l \qquad (2)$$

The Lagrange multipliers $\alpha_i \in (0, C)$ are generated from Equation (2), and the classification decision function (x) is then built as in Equation (3). $f(x) = \text{sign}(\sum_{i=1}^{l} \alpha_i y_i \left(\varphi(x_i).\varphi(x)\right) + b)$ (3)

The kernel function is often defined as K(xi, x) = $(\varphi(x_i).\varphi(x))$. SVM uses a variety of kernel functions for its classification purpose. However, the Gaussian kernel function is the one that is most frequently utilized in research studies. Because it does classification more effectively than the other SVM kernel functions, due to its approximation capabilities, the Gaussian kernel function was chosen for this work. Its form is expressed as follows:

$$k(x_i, y_j) = \exp\left(-\gamma \left\| x_i - x_j \right\|^2\right) \qquad (4)$$

Where γ is the kernel parameter within the Gaussian kernel function in Equation (4), typically, the terms "kernel parameter γ, and c, penalty parameter refer to SVM parameters with a Gaussian kernel function that the user should optimize.

2.3 Proposed ResNet 50_SVM Model

The ResNet-50 network model is one of the most effective models out of all of these models. There are 50 layers in the ResNet-50 model. One fully linked layer is employed at the network's end, while 49 of the 50 levels are convolution layers. There were 3.9 million MACs and 25.5 million weights in the entire network. The detailed architecture of the proposed model is discussed below.

2.3.1 Architecture

This section explains the architecture of the proposed ResNet-50-based SVM models. ResNet-50 is a deep learning algorithm using the dropout approach during training. Our suggested system was customized by replacing the ResNet-50's trainable classifier with an SVM classifier.

The major goal of this study is to enhance the suggested model's performance and create a new, effective recognition system motivated by the two formalisms. Figure 1 shows the detailed ResNet-50 model design. Additionally, Figure 2 shows the ResNet-50-based-SVM model's suggested network design.

3.3.2 Modelling

ResNet-50 normally accepts an input image with three color channels (red, green, and blue) and a resolution of 224x224 pixels. It belongs to the ResNet (Residual Network) family of models, which are renowned for effectively training extremely deep neural

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 43 Issue: 08-2024
DOI: 10.5281/zenodo.13292340

networks while minimizing the vanishing gradient issue. The second layer is a 3x3 max-pooling layer with a stride of 2, which is followed by a 7x7 convolutional layer with 64 filters as the first layer. Out of the 50 layers, 49 are convolution layers, while the last layer of the network uses one fully linked layer, which accepts input from the merged maps. As a result, they can extract characteristics that are increasingly resilient to local picture alterations. FCL is the 50th layer and has ten neurons in it.

We build a model by training the ResNet-50 network on the bird picture dataset. Input pictures pass via 49 convolutional layers during training. In order to create convolutional layers, 7x7x3-inch filters are used. Each of the convolution layers has a different number of filters. The detailed architecture of the ResNet-50 model is given in Table 1. SVM replaced the last layer for classification with an RBF kernel. Over-fitting can happen due to using a large amount of data and parameters.

Dropout is thus used to fix and avoid this issue on our network. Only the FCL layer, and more specifically, feed-forward connections, is affected by dropout (perceptron). This decision was made since over-fitting is not a concern because of the convolutional layers' limited number of parameters, and dropout would not have a significant impact. The SVM takes the outputs from the hidden units as a feature vector for the training process. After that, the training stage continues till realizing good training. Finally, the SVM classifier classified the test set with automatically extracted features.

Table 1: The detailed architecture of the ResNet-50 model

Layer	Kernel Size	Stride	Padding	Output Size
Input				[224 × 224 × 3]
Conv1	7 × 7 × 3	2	3	[112 × 112 × 64]
Max pool	3 × 3	2	-	[56 × 56]
Conv2	[1×1conv,64],[3 × 3conv,64],1 × 1conv,256]	2	-	[56 × 56]
	[1×1conv,64],[3 × 3conv,64],1 × 1conv,256]	1	-	
	[1×1conv,64],[3 × 3conv,64],1 × 1conv,256]	1	-	
Conv3	[1×1conv,128],[3 × 3conv,128],[1 × 1conv,512]	2	-	[28 × 28]
	[1×1conv,128],[3 × 3conv,128],[1 × 1conv,512]	1	-	
	[1×1conv,128],[3 × 3conv,128],[1 × 1conv,512]	1	-	
	[1×1conv,128],[3 × 3conv,128],[1 × 1conv,512]	1	-	
Conv4	[1×1conv,256],[3 × 3conv,256],[1 × 1conv,1024]	2	-	[14 × 14]
	[1×1conv,256],[3 × 3conv,256],[1 × 1conv,1024]	1	-	
	[1×1conv,256],[3 × 3conv,256],[1 × 1conv,1024]	1	-	
	[1×1conv,256],[3 × 3conv,256],[1 × 1conv,1024]	1	-	
	[1×1conv,256],[3 × 3conv,256],[1 × 1conv,1024]	1	-	
	[1×1conv,256],[3 × 3conv,256],[1 × 1conv,1024]	1	-	
Conv5	[1×1conv,512],[3 × 3conv,512],[1 × 1conv,2048]	2	-	[7 × 7]
	[1×1conv,512],[3 × 3conv,512],[1 × 1conv,2048]	1	-	
	[1×1conv,512],[3 × 3conv,512],[1 × 1conv,2048]	1	-	
Average pool	7 × 7	7	-	[1 × 1]
fc1000 softmax				1000

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 43 Issue: 08-2024
DOI: 10.5281/zenodo.13292340

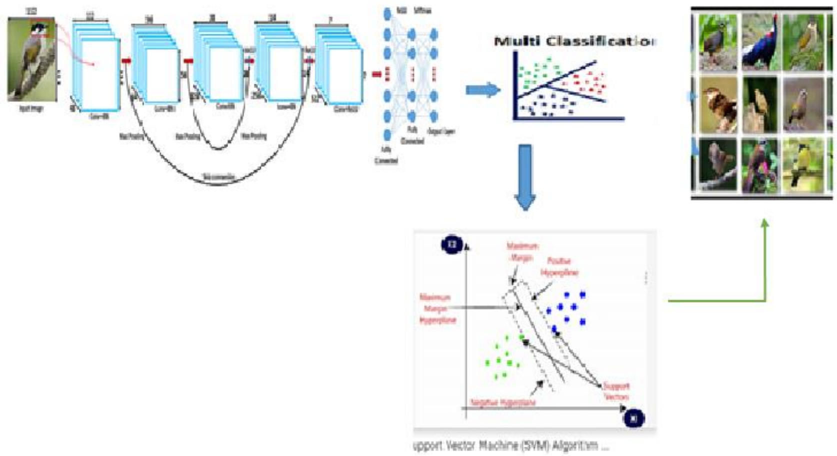

Figure. 2: The proposed ResNet50_SVM Model

In the proposed ResNet-50 model, the SVM has many advantages. One of the advantages of the proposed methods is that a support vector machine (SVM) can easily be integrated into a deep learning architecture. The trained SVM model retrieves feature vectors from the last fully connected deep learning model and provides class scores. Another advantage is that using SVM instead of the SoftMax classifier allows better classification accuracy for bird recognition. However, the proposed methods also have certain disadvantages. In conjunction with SVM models, deep learning models require two training tasks. First, deep learning models are trained, and then the models are created.

2.4 Dataset

This study used birds images datasets containing a total of 20000 images of ten different classes, the bird's images which consist of ten different classes such as ALBATROSS", "BOBOLINK," "CANARY," "CASSOWARY," "FLAMINGO," "GUINEAFOWL," "GYRFALCON," "HAWFINCH," "IBISBILL," "KIWI" mention in figure 3 below. In this research, 70 % of the dataset is used for training purposes, while the rest is used for testing the model. All the images are taken from the link given: https://www.kaggle.com/datasets/gpiosenka/100-bird-species/code, which consists of 100 different spices. This study selected ten different classes from this dataset.

Figure 3: Samples of the images in the Wild-Birds Dataset,

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 43 Issue: 08-2024
DOI: 10.5281/zenodo.13292340

2.5 Measuring Criteria

Various performance parameters such as accuracy, precision, recall, and f-score [30] are used in this paper. Precision is computed as:

$$Prec = \frac{TP}{TP+FP} \quad (1)$$

While TP stands for true positive rate and FP accurately displays false positive rate.

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

While in the recall, TP stands for the true positive rate and FN for the false negative rate.

$$MacroAvg\ F1 = \frac{\sum_{j=i}^{c} \frac{2*Prec_j*Recall_j}{Prec_j+Recall_j}}{C} \quad (3)$$

While Prec stands for the jth class's precision and Recall for its recall. The number of classes is multiplied by the macro-averaged decision of all classes.

$$Micro\ Avg\ F1 = \frac{2*Prec*Recall}{Prec+Recall} \quad (4)$$

Whereas $Prec$ denotes the precision and $Recall$ denotes the recall of the jth class. The decision of all classes is calculated in micro-averaged.

$$Acc = \frac{TP+TN}{TP+FP+TN+FN} \quad (5)$$

3. RESULT AND DISCUSSIONS

This section explains the overall result of the proposed model as compared to a similar hybrid deep learning model with SVM as a base classifier. A total of deep learning-based SVM models are simulated in this paper for different bird recognition. The performance of the used models is checked in terms of accuracy, recall, precision, and f-score. The detailed evaluation performance of the used models is in the table below. The six models that are employed in the simulation are listed below.

1. Proposed ResNet-50+SVM.
2. VGG_16+SVM
3. VGG_19+SVM
4. ResNet101 + SVM
5. DensNet121 + SVM
6. Mobile Net+ SVM

A thorough evaluation of model performance is given in Table 2, which includes precision, recall, F1 score, Cohen Kappa score, Matthews Correlation Coefficient, and accuracy. In particular, the ResNet-50 + SVM model that is suggested performs better than the other

ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 43 Issue: 08-2024
DOI: 10.5281/zenodo.13292340

models presented in this work. A precise precision score of 0.983, recall of 0.98, F1 score of 0.979, Cohen Kappa score of 0.977, Matthews Correlation Coefficient of 0.978, and accuracy of 0.98 are attained by the ResNet-50 + SVM model. A precision score of 0.967, recall of 0.976, F1 score of 0.965, Cohen Kappa score of 0.976, Matthews Correlation Coefficient of 0.961, and accuracy of 0.96 are achieved by the VGG_16 + SVM model, in contrast. Similar to the previous model, the VGG_19 + SVM model has a precision score of 0.976 and Recall was 0.978, F1 was 0.967, Cohen Kappa was 0.976, Matthews Correlation Coefficient was 0.966, and accuracy was 0.97. The ResNet101 + SVM model also obtains accuracy of 0.96, recall of 0.96, F1 score of 0.958, Cohen Kappa score of 0.9556, and scores of 0.9573, 0.9573, and 0.9714 for precision. In a similar vein, the DenseNet121 + SVM model yields precision, recall, F1 score, Cohen Kappa, Matthews Correlation Coefficient, and accuracy scores of 0.96, 0.96, 0.9556, 0.9536, and 0.96 respectively. Finally, the MobileNet + SVM model attains a precision score of 0.977, a recall of 0.965, an F1 score of 0.965, a Cohen Kappa score of 0.956, a Matthews Correlation Coefficient of 0.961, and an accuracy of 0.96. Figures 4, 5, and 6 provide graphical representations of these performance evaluation parameters, comparing the utilized models to the proposed ResNet-50 + SVM model discussed in this paper.

Table 2: Performance Evaluation of the Deep Learning-based SVM Models

Algorithms	Precision	Recall	F1 Score	Cohen Kappa Score	Matthews Corrcoef	Accuracy
ResNet50 + SVM	0.983	0.98	0.979	0.977	0.978	0.98
VGG_16 + SVM	0.967	0.976	0.965	0.976	0.961	0.96
VGG_19 + SVM	0.976	0.978	0.967	0.976	0.966	0.97
ResNet101 + SVM	0.9714	0.96	0.958	0.9556	0.9573	0.96
DensNet121 + SVM	0.96	0.96	0.96	0.9556	0.9536	0.96
Mobile Net+ SVM	0.977	0.965	0.965	0.956	0.961	0.96

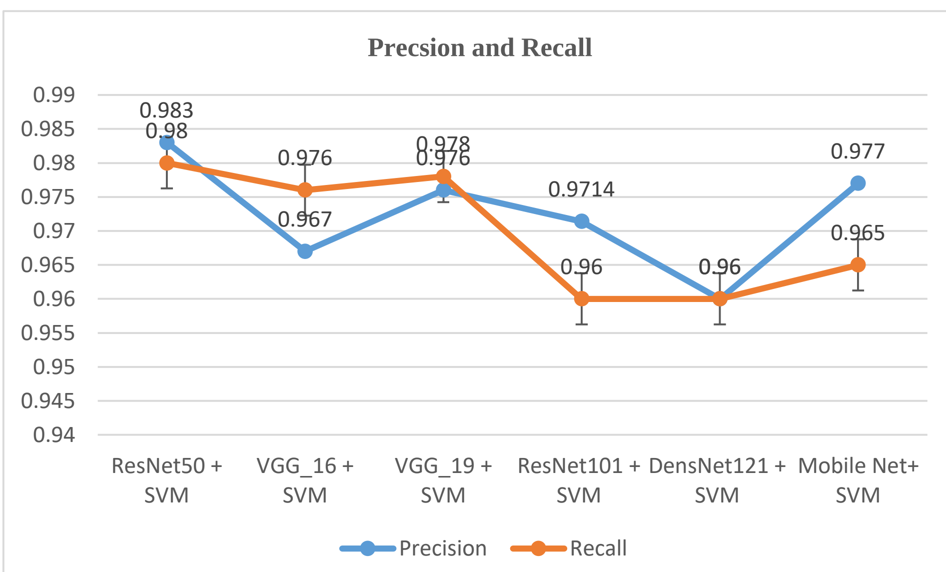

Figure 4: Precision and Recall Performances

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 43 Issue: 08-2024
DOI: 10.5281/zenodo.13292340

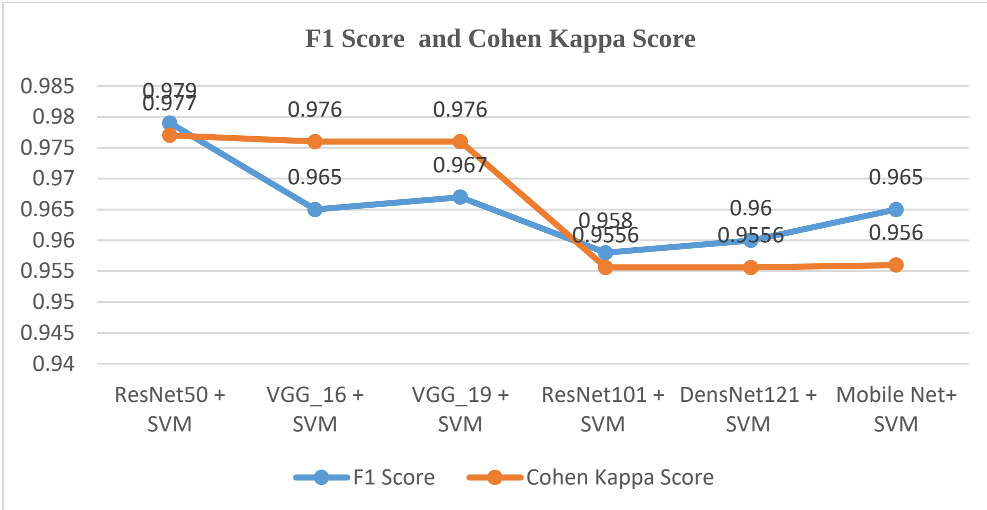

Figure 5: F1 Score and Cohen Kapa Score Performances

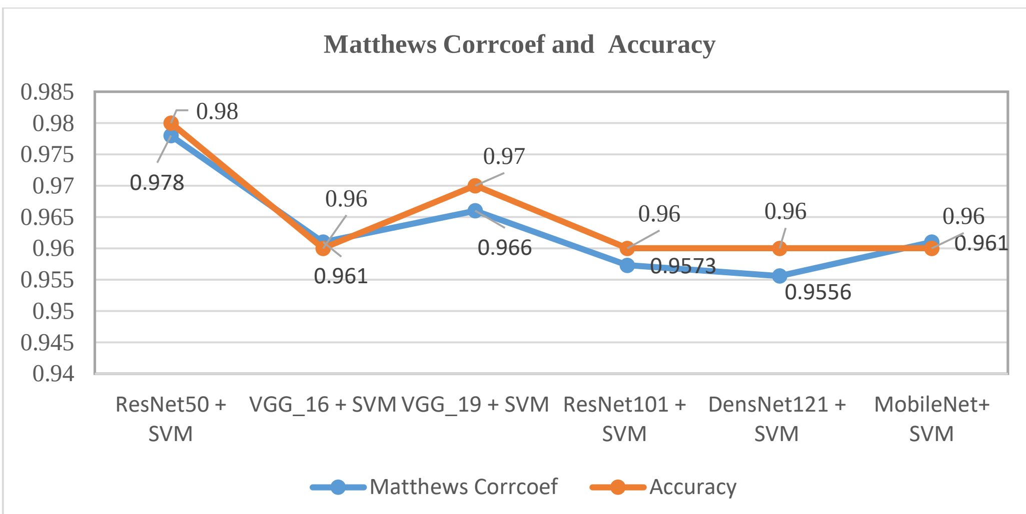

Figure 6: Matthews Corrcoef and Accuracy Performances

Likewise, Tables 3, 4, and 5 provide an extensive assessment of the performance of the proposed models in comparison to the models used in this research. Table 3 specifically outlines the class-specific performance evaluation of the ResNet-50 + SVM and VGG-16 + SVM models, including precision (P), recall (R), F1-score (F1), macro-average (W_avg), and weighted-average (W_avg) metrics. In terms of the overall results presented in the table, it becomes evident that the proposed model exhibits superior performance when contrasted with the other models. The dataset employed in this study also contains 20,000 bird photos spread over ten different classes, according to a class-wise split. For instance, in the case of class 0, denoted as "ALBATROSS," the suggested model achieves a P score of 1, R score of 1, and an F1 score of 1. In contrast, the VGG-16 + SVM model achieves an F1-score of 0.978, P score of 0.967, and R scores of 0.967 for the same class. Similar results are shown by the VGG-19 + SVM model, which has an

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 43 Issue: 08-2024
DOI: 10.5281/zenodo.13292340

F1-score of 0.968, P score of 0.976, and R score of 0.945. The ResNet101 + SVM model also achieves an F1-score of 1 for class 0, a P score of 1, and a recall score of 1. Whereas the DenseNet121 + SVM model gets an F1-score of 1, a P score and R of 1, and both. In contrast, the MobileNet + SVM model records for the same class a P score of 0.731, a R score of 0.801, and an F1 score of 0.782.

Table 3: ResNet50 + SVM and VGG_16 + SVM Classification Report

ResNet50 + SVM				VGG_16 + SVM			
Classes	**Precision**	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-score**	**Support**
0	1	1	1	0.967	0.967	0.978	2000
1	1	0.8	0.888	0.966	0.976	0.966	2000
2	0.833	1	0.909	0.956	0.954	0.966	2000
3	1	1	1	0.955	0.965	0.966	2000
4	1	1	1	0.967	0.965	0.976	2000
5	1	1	1	0.966	0.976	0.954	2000
6	1	1	1	0.966	0.954	0.965	2000
7	1	1	1	0.966	0.965	0.976	2000
8	1	1	1	0.966	0.976	0.954	2000
9	1	1	1	0.967	0.966	0.977	2000
Accuracy			0.98			0.96	20000
Macro avg	0.983	0.98	0.979	0.966	0.975	0.976	20000
Weight avg	0.983	0.98	0.979	0.966	0.975	0.976	20000

The suggested model has a P score of 1, R of 0.8, and an F1 score of 0.888 for class 1, designated as "BOBOLINK." For the same class, the VGG_16 + SVM model performs better, with P scores of 0.966, R scores of 0.976, and F1 scores of 0.978. The P, R, and F1 scores for the VGG_19 + SVM model are identical, coming in at 0.966, 0.975, and 0.975.

Additionally, the DenseNet121 + SVM model achieves P, and R scores of 1, as well as an F1 score of 1, whereas the ResNet101 + SVM model earns a P of 1, R of 0.6, and an F1 score of 0.75 for class 1. The MobileNet + SVM model, on the other hand, achieves a P score of 0.7343, R of 0.6, and an F1 score of 0.75 for class 1.

For class 2, often known as "CANARY," the suggested model achieves a P score of 0.833, R score of 1, and an F1 score of 0.909. For the same class, the VGG_16 + SVM model records a P score of 0.956, R of 0.954, and an F1score of 0.966. The VGG_19 + SVM model exhibits a P of 0.965, R of 0.967, and F1 of 0.973.

Additionally, the ResNet101 + SVM model earns a P of 0.714, R score of 1, and an F1 score of 0.833 for class 2, whereas the DenseNet121 + SVM model achieves P and R scores of 1, as well as an F1score of 1. On the other hand, the MobileNet + SVM model records P score of 0.714, R of 0.723, and an F1score of 0.833 for class 2.

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 43 Issue: 08-2024
DOI: 10.5281/zenodo.13292340

Table 4: VGG_19 + SVM, and ResNet101 + SVM Classification Report

VGG_19 + SVM				ResNet101 + SVM			
Classes	**Precision**	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-Score**	**Support**
0	0.976	0.945	0.968	1	1	1	2000
1	0.966	0.975	0.975	1	0.6	0.75	2000
2	0.965	0.967	0.973	0.714	1	0.833	2000
3	0.954	0.986	0.971	1	1	1	2000
4	0.963	0.958	0.947	1	1	1	2000
5	0.976	0.945	0.968	0.963	0.977	0.975	2000
6	0.963	0.977	0.975	0.962	0.965	0.974	2000
7	0.962	0.965	0.974	0.957	0.986	0.973	2000
8	0.957	0.986	0.973	0.963	0.958	0.947	2000
9	0.963	0.958	0.947	0.963	0.977	0.975	2000
Accuracy			0.97			0.96	20000
Macro avg	0.966	0.963	0.972	0.971429	0.96	0.958333	20000
Weight avg	0.965	0.956	0.968	0.971429	0.96	0.958333	20000

The suggested model displays P of 1, R of 1, and F1 score values of 1 for class 3, often known as "CASSOWARY." The VGG_16 + SVM model, in contrast, achieves P value of 0.955, R value of 0.965, and an F1 score of 0.966 for the same class. The P, R, and F1 scores for the VGG_19 + SVM model are all 0.954, 0.986, and 0.971, respectively. Additionally, for class 3, the ResNet101 + SVM model achieves P of 1, R of 1, and F1score values of 1, whereas the DenseNet121 + SVM model only records P, R, and F1values of 0.8, 0.8, and 0.8. For class 3, the MobileNet + SVM model, on the other hand, records P, values of 0.745, R values of 0.701, and an F1 score of 0.812.

For class 4, called "FLAMINGO," the suggested model exhibits P of 1, R of 1, and F1 score values of 1. For class 4, the VGG_16 + SVM model, in contrast, achieves P values of 0.967, R values of 0.965, and an F1 score of 0.976. The VGG_19 + SVM model performs similarly, scoring 0.963 for P, 0.958 for R, and 0.947 for F1. Additionally, for class 4, the ResNet101 + SVM model achieves P, R of 1, and an F1 score of 1, but the DenseNet121 + SVM model only records P, R and F1 values of 0.8, 0.8, and 0.8. For class 4, the MobileNet + SVM model, on the other hand, records P, value of 1, R of 1, and F1 score of 1. For class 5, known as "GUINEAFOWL," the suggested model achieves P of 1, R of 1, and F1 score values of 1. For the same class, the VGG_16 + SVM model achieves P values of 0.966, R values of 0.976, and an F1 score of 0.954. Similar results are shown by the VGG_19 + SVM model, which has an F1 score of 0.968, P score of 0.976, and R score of 0.945. Additionally, the DenseNet121 + SVM model obtains P, R of 1, and F1 score values of 1, while the ResNet101 + SVM model achieves P, R, and F1 score values of 0.963, 0.977, and 0.975 for class 5. On the other hand, for class 5, the MobileNet + SVM model records P value of 1, R of 1, and F1 score of 1.

For class 6, known as "GYRFALCON," the suggested model achieves P of 1, R of 1, and F1 score values of 1. In contrast, the VGG_16 + SVM model reports for the same class

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 43 Issue: 08-2024
DOI: 10.5281/zenodo.13292340

P value of 0.966, R value of 0.954, and an F1 score of 0.965. The P, R, and F1 scores for the VGG_19 + SVM model are all 0.963, 0.977, and 0.975, respectively. Additionally, the ResNet101 + SVM model obtains P of 1, R of 1, and F1 score values of 0.962, 0.965, and 0.974 for class 6, compared to P, R, and F1 score values of 1 for the DenseNet121 + SVM model. On the other hand, for class 6, the MobileNet + SVM model records P value of 1, R of 1, and an F1 score of 1.

Table 5: DensNet121 + SVM, Mobile Net+ SVM Classification Report

DensNet121 + SVM				Mobile Net+ SVM			
Classes	**Precision**	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-Score**	**Support**
0	1	1	1	0.731	0.801	0.782	2000
1	1	1	1	0.743	0.6	0.75	2000
2	1	1	1	0.714	0.723	0.833	2000
3	0.8	0.8	0.8	0.745	0.701	0.812	2000
4	0.8	0.8	0.8	1	1	1	2000
5	1	1	1	1	1	1	2000
6	1	1	1	1	1	1	2000
7	1	1	1	0.731	0.801	0.782	2000
8	1	1	1	0.742	0.6	0.75	2000
9	1	1	1	0.714	0.723	0.833	2000
Accuracy			0.96			0.96	20000
Macro avg	0.96	0.96	0.96	0.82	0.845	0.843	20000
Weight avg	0.96	0.96	0.96	0.834	0.852	0.843	20000

The suggested model displays P of 1, R of 1, and F1 score values of 1 for class 7, often known as "HAWFINCH." In contrast, the VGG_16 + SVM model reports class 7 P and R values of 0.966, 0.965, and 0.976 for the F1 score. The P, R, and F1 scores for the VGG_19 + SVM model are all 0.962, 0.965, and 0.974, respectively. Additionally, the DenseNet121 + SVM model obtains P of 1, R, and F1 score values of 1 compared to the ResNet101 + SVM model's 0.957, 0.986, and 0.973 scores for class 7. On the other hand, the MobileNet + SVM model records for class 7 P value of 0.731, R value of 0.801, and an F1 score of 0.782. For class 8, referred to as "IBISBILL," the suggested model exhibits P of 1, R of 1, and F1 score values of 1. The P value, R value, and F1 score for the same class are all recorded as 0.766, 0.776, and 0.754 by the VGG_16 + SVM model, respectively. Similar results are shown by the VGG_19 + SVM model, which has an F1 score of 0.947, P score of 0.957, and R score of 0.958. Additionally, the DenseNet121 + SVM model obtains P of 1, R, and F1 score values of 1, whereas the ResNet101 + SVM model achieves P, R, and F1 score values of 0.957, 0.958, and 0.947 for class 8. For class 8, the MobileNet + SVM model, on the other hand, records P values of 0.742, R values of 0.6, and F1 scores of 0.75. For class 9, known as "KIWI," the suggested model shows P of 1, R of 1, and F1 score values of 1. For the same class, the VGG_16 + SVM model achieves P values of 0.967, R values of 0.966, and F1 score values of 0.977. The P, R, and F1 scores for the VGG_19 + SVM model are all 0.963, 0.958, and 0.947, respectively. Additionally, the DenseNet121 + SVM model obtains P, R, and F1 score values of 1, while the ResNet101 + SVM model achieves P, R, of 1 and F1 score values

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 43 Issue: 08-2024
DOI: 10.5281/zenodo.13292340

of 0.963, 0.977, and 0.975 for class 9. In contrast, the MobileNet + SVM model records class 9 P values of 0.714, 0.723 R, and 0.833 F1 score.

4. CONCLUSION

This study presents a deep learning-based ResNet50 model, trained with support vector machines as a based classifier. The SVM classifier is also combined in this study with some selected deep learning-based models such as VGG_16, VGG_19, ResnNet101, DensNet121, and Mobil Net. Using images of birds, the performance of the employed models is evaluated. Support vector machines are utilized to categorize the appearances of birds in images after a ResNet50 architecture is developed to simulate their looks. Instead of using the SoftMax approach, this technique supports vector machines for a more reliable and accurate categorization of bird images. Then, another deep learning approach was created based on support vector machines, including VGG_16, VGG_19, ResnNet101, DensNet121, and Mobil Net. This method, similar to the one before it, extracts attributes from photographs before classifying them using support vector machines. The suggested approaches use SVM in place of the SoftMax layer. The suggested approaches offer improved accuracy compared to conventional methods because support vector machines are more potent classifiers than SoftMax. The performance findings demonstrate that the suggested models outperform the commonly used deep learning models, such as the Mobil Net-based SVM models and the VGG_16, VGG _19, ResnNet101, and DensNet121.

Funding statement: The authors did not receive any funding for this work.

Declarations Conflict of Interests: The author of this paper declares that there is no conflict of interest regarding the publication of this paper.

References

1) Favorskaya, M. and AJPCS Pakhirka, *Animal species recognition in the wildlife based on muzzle and shape features using joint CNN.* 2019. **159**: p. 933-942.

2) Linchant, J., et al., *Are unmanned aircraft systems (UAS s) the future of wildlife monitoring? A review of accomplishments and challenges.* 2015. **45**(4): p. 239-252.

3) Gilmer, D.S., et al., *Goose counts from aerial photographs using an optical digitizer.* 1988. **16**(2): p. 204-206.

4) Cunningham, D.J., W.H. Anderson, and RMJWSB Anthony, *An image-processing program for automated counting.* 1996. **24**(2): p. 345-346.

5) Trathan, PNJWSB, *Image analysis of color aerial photography to estimate penguin population size.* 2004. **32**(2): p. 332-343.

6) Abd-Elrahman, A., et al., *Development of pattern recognition algorithm for automatic bird detection from unmanned aerial vehicle imagery.* 2005. **65**(1): p. 37.

7) Liu, C.-C., Y.-H. Chen, and H.-L.J.E.i. Wen, *Supporting the annual international black-faced spoonbill census with a low-cost unmanned aerial vehicle.* 2015. **30**: p. 170-178.

8) Hong, S.-J., et al., *Application of deep-learning methods to bird detection using unmanned aerial vehicle imagery.* 2019. **19**(7): p. 1651.

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 43 Issue: 08-2024
DOI: 10.5281/zenodo.13292340

9) Ammour, N., et al., *Deep learning approach for car detection in UAV imagery.* 2017. **9**(4): p. 312.

10) Krizhevsky, A., I. Sutskever, and G.E.J.A.i.n.i.p.s. Hinton, *Imagenet classification with deep convolutional neural networks.* 2012. **25**.

11) Girshick, R., et al. *Rich feature hierarchies for accurate object detection and semantic segmentation*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014.

12) Girshick, R. *Fast r-cnn*. in *Proceedings of the IEEE International Conference on Computer Vision*. 2015.

13) Ren, S., et al., *Faster r-cnn: Towards real-time object detection with region proposal networks.* 2015. **28**.

14) Redmon, J., et al. *You only look once: Unified, real-time object detection*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

15) Liu, W., et al. *Ssd: Single shot multibox detector*. in *European conference on computer vision*. 2016. Springer.

16) Lin, T.-Y., et al. *Focal loss for dense object detection*. in *Proceedings of the IEEE International Conference on Computer Vision*. 2017.

17) Szegedy, C., et al. *Going deeper with convolutions*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.

18) andola, F.N., et al., *SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size.* 2016.

19) Faster, R.J.A.i.n.i.p.s., *Towards real-time object detection with region proposal networks.* 2015. **9199**(10.5555): p. 2969239.2969250.

20) Simonyan, K. and A.J.a.p.a. Zisserman, *Very deep convolutional networks for large-scale image recognition.* 2014.

21) He, K., et al. *Deep residual learning for image recognition*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

22) Huang, G., et al. *Densely connected convolutional networks*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

23) Chang, Y.-C., et al. *Pedestrian detection in aerial images using vanishing point transformation and deep learning*. in *2018 25th IEEE International Conference on Image Processing (ICIP)*. 2018. IEEE.

24) Chen, F., et al., *Fast automatic airport detection in remote sensing images using convolutional neural networks.* 2018. **10**(3): p. 443.

25) Maire, F., L.M. Alvarez, and A. Hodgson. *Automating marine mammal detection in aerial images captured during wildlife surveys: a deep learning approach*. in *Australasian Joint Conference on Artificial Intelligence*. 2015. Springer.

26) Guirado, E., et al., *Automatic whale counting in satellite images with deep learning.* 2018: p. 443671.

27) Wu, S., et al., *Deep residual learning for image steganalysis.* 2018. **77**(9): p. 10437-10453.

28) Vapnik, V., *The nature of statistical learning theory Springer New York Google Scholar.* 1995.

29) Fernandez-Lozano, C., et al., *Hybrid model based on genetic algorithms and SVM applied to variable selection within fruit juice classification.* The Scientific World Journal, 2013. **2013**.

30) Tang, L. and H. Liu. *Bias analysis in text classification for highly skewed data*. in *Fifth IEEE International Conference on Data Mining (ICDM'05)*. 2005. IEEE.