International Journal on Advanced Science Engineering Information Technology

Vol.13 (2023) No. 5
ISSN: 2088-5334

Bird Species Recognition System with Fine-Tuned Model

Ching-Yang Ngo, Lee-Ying Chong*, Siew-Chin Chong, Pey-Yun Goh

Faculty of Information Science and Technology, Multimedia University, Melaka, Malaysia
*Corresponding author: *lychong@mmu.edu.my*

***Abstract*—A bird recognition system identifies bird species by combining computer vision and machine learning techniques to categorize different bird species with high accuracy. Moreover, the bird species recognition system represents a significant advance in animal protection and zoological research, especially for the rare and elusive bird species living in the jungle. This work focuses on an image-based system for bird species recognition. In bird species recognition, users input the bird images, and the system uses a deep learning model trained for optimal results in identifying different bird species from the images. We used fine-tuned deep learning models (Inception-V3 and EfficientNet-B4) to evaluate and determine which model can best perform image-based bird species recognition. Several unique datasets were used to evaluate and determine which model was best suited for image-based bird species recognition. These datasets consist of CUB -200-2011, Kaggle-510 bird species, 325 bird species, and a self-generated dataset (100 bird species from Malaysia). When applied to these four different datasets, the experimental results clearly show the advantage of fine-tuning the deep learning models. This study makes an important contribution to ornithology by providing a robust and trustworthy method for identifying and cataloging bird species, especially those that are rarely seen in the wild. Thus, the bird identification system is important for scientific research and animal welfare.**

***Keywords*— Bird species recognition system; fine-tuned model; machine learning; deep learning.**

Manuscript received 18 Oct. 2022; revised 7 Mar. 2023; accepted 21 Jun. 2023. Date of publication 31 Oct. 2023.

I. INTRODUCTION

The advent of artificial intelligence has changed the lifestyle of people, and it further enhanced the life quality of the people [1]–[3]. It enables individuals to alleviate their workload by imbuing machines with human-like capabilities and even surpassing human thinking abilities. For example, artificial technologies enable the identification of rare animal species, distinguishing between various individuals' voices, and much more.

The world's living species are impacted by environmental changes and climatic circumstances brought on by science and technical progress [4]–[6]. The most severely affected group is the birds, mainly living in the woods. Therefore, protecting small or rare bird species to prevent them from extinction is necessary [7]–[9]. To do this, classifying the bird species is the first and most crucial step, which helps animal protection experts and zoologists to know the different species of birds that live in the forest, especially the rare species, to advocate the conservation of rare birds and their habitats.

Convolutional neural networks (CNNs) architecture exhibits a superior design for most computer vision applications [10]–[12]. The deep learning approach has the capability to perform the learning of high-level features from data gradually. Hence, the deep learning approach outperforms machine learning techniques in many domains when a large amount of data is provided [13], [14]. The present research only utilizes some bird species in the dataset for recognition such as in [15] and [16] resulting in an inability to accurately demonstrate its performance in recognition.

Qiao et al. [15] proposed a Support Vector Machine (SVM) classifier mixed with a decision tree for bird recognition using feature extraction based on beak width and eye-beak root range shift. The decision tree grouped similar birds into a set of birds of the same theme. The classifier achieved 83.37% accuracy by incorporating all features such as color, beak, and head color for both SVM and decision tree.

Gavali et al. [16] proposed using a deep convolutional neural network (DCNN) for bird species recognition with the CUB-200-2011 dataset. They pre-processed the data by creating a grayscale image dataset and reducing the image size. The DCNN consisted of different layers of convolutional neural networks, and the GoogLeNet framework was used for recognition. The dataset was retrained to maximize classification accuracy. The input was compared with the trained dataset to classify the species, and the system achieved an accuracy of 88.33%.

Jain et al. [17] presented using Convolutional Neural Networks (CNN) for bird species recognition using bird audio. They extracted acoustic features from bird sound files using Mel Frequency Cepstral Coefficients (MFCC) and created a dataset with these features and classes. Keras sequential model was used to predict the label belonging to the bird from the audio file. The model achieved 70% accuracy in detecting bird species, as confirmed by the evaluation of their model.

Sharma et al. [18] used two pre-trained neural networks, ResNet50V2 and EfficientNetB0, for video-based bird species recognition with a dataset of 137 bird species. They extracted and pre-processed the image and audio from the input video, fed them to the respective networks, and obtained two outputs: image classification and audio classification. These outputs were compared to the dataset to determine the common bird species in the input video, achieving an accuracy score of 90%.

Wang et al. [19] developed a bird species recognition system based on sound using Recurrent Neural Networks (RNN). They collected a dataset of 72,172 bird audio samples of 264 species and performed pre-processing steps such as high pass filtering and noise removal. The inputs to the RNN model were MFCCs and Mel-spectrograms, and LSTM was used as the baseline network to improve classification performance. They achieved an accuracy score of 74.94%.

Pahuja et al. [20] developed a Multi-layer Perceptron Artificial Neural Network (MLP-NN) for recognizing bird species by analyzing bird sound spectrograms. They extracted Mean Instantaneous Frequency (MIF), Mean Instantaneous Bandwidth (MIB), and Group Delay (GD) features from the spectrum of eight different bird species. These features were then fed into the MLP-NN classifier, which is a multi-class classifier for layered neural networks trained using the feed-forward and backpropagation approach. The researchers conducted several experiments to optimize the performance of the MLP-NN, such as changing the number of neurons in the hidden layers, epoch size, and learning rate. The best accuracy of 96.1% was achieved with 300 epochs, a batch size of nine, and 24 neurons in the hidden layers.

Ragib et al. [21] proposed a deep-learning model and pre-trained ResNet for image-based bird species recognition. They used both a model with randomly assigned weights and a pre-trained ResNet18 model, which was trained on the ImageNet dataset. The images were resized to 224x224x3 before fitting into the model. The ResNet18 model achieved a top-5 accuracy of 96.71%, while the random weight model achieved 63.48%. The authors also evaluated the accuracy of different ResNet models and found that ResNet101 achieved the highest accuracy of 97.98%.

Alswaitti et al. [22] evaluated bird recognition systems using three different approaches: traditional machine learning classifiers, CNNs, and transfer learning based CNNs. They found that traditional machine learning classifiers had lower accuracy when more bird species were added, ranging from 6% to 50% accuracy for 20 species. CNNs like AlexNet, DenseNet-121, GoogLeNet, and ResNet-50 had better accuracy than traditional machine learning classifiers, but accuracy still dropped with more bird species added. Transfer learning based CNNs had the best accuracy, reaching up to 98%. The authors suggested that transfer learning is effective for deep learning with limited data sets, as it allows the transfer of model parameters for image recognition.

Jha et al. [23] developed a bird species recognition platform based on sequential neural networks and image processing. To improve classification accuracy, they trained and retrained a dataset, considering factors like beak shape and appearance. The system used a deep convolutional network to extract features and a pre-trained dataset for comparison. The workflow involved applying sequential-based CNN, extracting features, comparing the input to trained data, and recognizing the bird species. Their model achieved 84.76% accuracy.

Gómez-Gómez et al. [24] explored the performance of small-footprint deep neural networks (MobileNetV2) in classifying bird species based on audio and compared it with VGG16 and ResNet50. They collected 20 bird audio and used spectrograms generated from 1-second windows for feature extraction. Fine-tuning was done by freezing all layers in the network's body and training only the new fully connected head as a warm-up phase. The results showed that MobileNetV2 had similar or better accuracy than larger models.

Kondaveeti et al. [25] proposed an Arduino-based bird species recognition system for automatic bird species recognition. This system was based on the Arduino Uno, ESP-32 camera, and PIR Motion Sensor. The ESP-32 camera detected the motion in this system and uploaded the captured images to the drive. Then, the uploaded images fit into the trained deep-learning model. They used the ResNet101V2 CNN to build their classification model. They set the epochs to 5, and the batch size was 276, which took 22 hours for training. As a result, they obtained 87% validation accuracy in the proposed system.

Kumar et al. [26] experimented with a bird species recognition system using various deep learning networks such as MobileNet, AlexNet, InceptionResNet V2, InceptionV3, and EfficientNet. To conduct this experiment, they used a dataset from the Kaggle, which consists of 11488 images from 200 species. Besides, they used the data augmentation method to increase the bird images to 40000. In the augmentation technique, they performed re-scaling, cropping, zooming, rotating, etc. The experiment used 10 to 50 epochs for training accuracy and 50 and 100 epochs for testing accuracy. Experiment results showed MobileNet and EfficientNet were the quickest models, and EfficientNet achieved the highest accuracy, 98.25%.

Huang et al. [27] proposed a skip connection Convolutional Neural Network (CNN) to build a bird species recognition system which is able to recognise 27 bird species endemics to Taiwan. In their CNN model architecture, there were two fully connected layers and one softmax output layer. Besides, each convolutional layer used the 5x5 filter, and the batch normalisation layer included the ReLu activation function and the pooling layers. With the use of skip-connection, the model improved the feature extraction through weighted summation of the corresponding layers. As a result, their proposed skin-connection CNN model reached

99% accuracy compared to 93.98% of the typical CNN model.

This paper aims to examine and recognize bird species worldwide using deep learning approaches by involving all the bird species from the dataset. Besides, we also perform the model training in several new datasets such as Kaggle-510 Bird Species, 325-Bird Species, and self-collected 100-Malaysia-Birds dataset. Two deep learning algorithms, Inception-V3 and EfficientNet-B4, are employed to enable the system to classify bird species based on user input images. This paper analyses experiment results obtained from two algorithms in four bird datasets and explores the impact of fine-tuning a pre-trained CNN model.

II. Materials and Method

A. Bird Datasets

The dataset known as Caltech-UCSD Birds-200-2011, also called CUB-200-2011 dataset, holds the reputation of being extensively employed in bird recognition systems. The dataset comprises 11,788 images of 200 bird species, with approximately 60 images captured for each species. Each image captures a single bird species in a single frame. The images have different sizes, necessitating image pre-processing to achieve standardized image dimensions. While numerous papers have utilized this dataset, they have yet to formally report results involving all the bird species in the dataset. Consequently, this project aims to compute accurate results by considering all the bird species in the dataset when evaluating the performance.

Additionally, this paper utilizes two other bird datasets, namely the Kaggle 510 Bird Species Dataset and the 325 Bird Species Dataset. The Kaggle 510 Bird Species Dataset encompasses 510 bird species, including 81,950 training images, 2,550 testing images, and 2,550 validation images. The bird occupies more than 50% of the total pixels in each image. 325 Bird Species Dataset contains 50591 images from 325 different bird species. This dataset is split into 47341 training images, 1625 testing, and 1625 validation images. In the testing and validation dataset, there are five images in each of the species. Both datasets are suitable for fine-tuning the deep learning model as they contain more than 30,000 images, and the model can learn enough features of each species.

Data collecting is challenging in deep learning experiments, and this type of small dataset might make this experiment more accurate as the experiment can assess the different sizes of the datasets [28]. Hence, another bird dataset, the 100-Malaysia-Birds dataset, was created specifically for this study. This dataset comprises 100 bird species commonly found in Malaysia and was generated using the online crawl technique. This dataset contains 9,286 images, making it the smallest dataset used in this paper. The purpose of including this dataset is to assess the model's performance when working with a smaller dataset. Table I exhibits the characteristics of each of the datasets.

TABLE I
CHARACTERISTICS OF EACH DATASET

Dataset	No. of Species	Total Images
CUB-200-2011 Dataset	200	11788
Kaggle 510 Bird Species Dataset	510	86950
325 Bird Species Dataset	325	50591
100-Malaysia-Birds Dataset	100	9286

B. The Proposed Method

The proposed bird species recognition system consists of four phases: pre-processing, feature extraction, model training, and performance evaluation. Fig. 1 illustrates the proposed bird species recognition system.

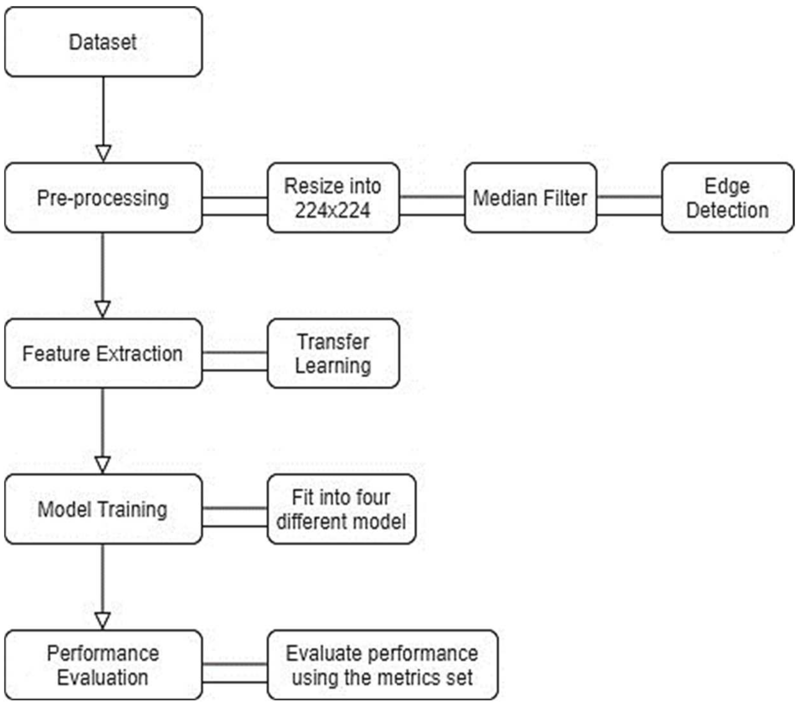

Fig. 1 The proposed bird species recognition system

C. Pre-processing

In the pre-processing stage, all of the images in each dataset are different, so all images are resized into 224 x 224 for a standard setup. The resized images are then normalized and standardized to have a range of [0, 1]. The median filter is performed to reduce the noise from the images. The median filter processes the image pixel by pixel, and each pixel compares itself with the neighborhood pixel and replaces it with the median value of both pixels. Then, edge detection is used to perform the image segmentation, and the Sobel operator is used to calculate the intensity gradient at each pixel to estimate the direction and rate of change in darkness from light.

D. Feature Extraction

The feature extraction is used to leverage two pre-trained models, which are Inception-V3 and EfficientNet-B4. In this case, transfer learning is used to utilize the features extracted during previous training on a large-scale dataset and apply them to a specific task. These models serve as feature extractors by capturing high-level features from the input images. Then, these features are convolutional layers of representations that have been learned. Subsequent layers can then utilize these representations to extract features for categorization. The equation of the transfer learning feature extraction is shown below:

$$Features = f\_base\_model\,(InputImage) \qquad (1)$$

in which *f_base_model* is the feature extraction function performed by either EfficientNet-B4 or Inception-V3, and *InputImage* represents the input image that needs to be processed. Applying transfer learning allows for more effective and efficient analysis of the particular dataset due to the deep learning models' capacity to grasp intricate patterns and features.

E. Inception-V3

Inception-V3, a widely recognized pre-trained deep-learning technique, is renowned for its effectiveness in image classification. Developed by the Google Research Team, it represents the third version of the Inception Network [29]. Inception-V3 incorporates numerous optimizations, including efficient grid size reduction, smaller convolutions, asymmetric components, and auxiliary classifiers. Compared to its predecessors, Inception-V3 boasts more layers, enhancing the performance and efficiency of image recognition tasks.

Fig. 2 presents the flow of the Inception-V3. In the Inception-V3, the GlobalAveragePooling2D layer is added to take the averages of the feature maps. Then, two fully connected layers are added to classify the image based on the output from the previous convolutional layers. The first connected layer is activated by rectified linear activation named ReLU [30]. ReLu is one of the popular activation functions in the deep learning approach that produces zero if the inputs are negative and the input value if it is positive, assisting gradient flow and encouraging sparse activation. The last fully connected layer, also an output layer of the Inception-V3 model, uses the SoftMax activation function.

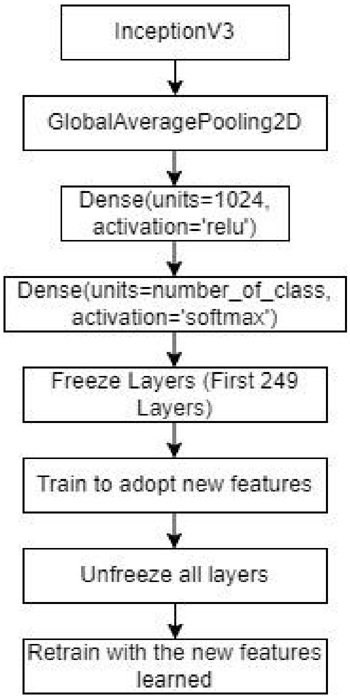

Fig. 2 The process flow of Inception-V3

SoftMax is an activation function for multiclass classification that converts input values into a probability distribution over several classes [31]. The next step is to freeze the first 249 layers from the Inception-V3 for training to adopt the new feature obtained from the dataset. Another reason for freezing the first 249 layers is that it yields optimal results during training. Lastly, all layers are unfrozen for retraining purposes with the new features obtained and evaluate the performance.

F. EfficientNet-B4

EfficientNet-B4, a convolutional neural network model, was introduced by Mingxing Tan and Quoc Le of Google Research [32]. It belongs to the EfficientNet model family, renowned for its high accuracy and computational efficiency. With its 157 layers, this model is particularly suitable for computationally intensive tasks that demand exceptional accuracy. The architecture employs depth-wise convolutions, squeeze-and-excitation modules, and other techniques to optimize the trade-off between size and accuracy.

EfficientNet-B4 utilizes a compound scaling method that adjusts the model's depth, width, and resolution to achieve a desirable balance. This model has showcased remarkable performance across various computer vision applications, including image classification, object detection, and segmentation. Same with the Inception-V3, the GlobalAveragePooling2D layer and two fully connected layers are added into EfficientNet-B4. However, the EfficientNet-B4 is a bigger model than the Inception-V3 model, so more layers are frozen for the train to adopt new features from the dataset. In this experiment, the adoption of new features involved freezing 469 layers, and it was found that this specific number of frozen layers produced optimal results. After that, all layers are unfrozen and retrained to evaluate the performance.

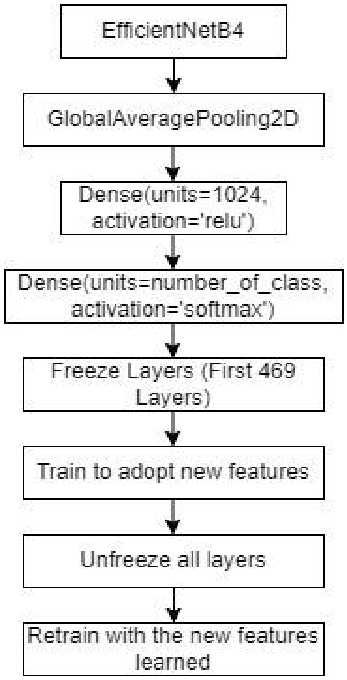

Fig. 3 The process flow of EfficientNet-B4

III. Result and Discussion

A. Experimental Measure

There are four metrics used to evaluate the performance of the proposed system. There is an accuracy score, precision score, recall score, F1-score, as listed in Table II.

TABLE II
PERFORMANCE METRICS

Metrics	Description
Accuracy score	It refers to the proportion of accurate predictions to all input samples. $Accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)}$ *TP* and *TN* represent the true positive and true negative while *FP* and *FN* represent the false positive and false negative.
Precision score	It refers to the proportion of the successfully identified positive position in the confusion matrix that was forecasted. $Precision = \frac{TP}{(TP + FP)}$
Recall score	It calculates the ratio of the projected favorable findings to all the actual label observations. $Recall = \frac{TP}{(TP + FN)}$
F1-score	It takes the precision score and recall score to calculate the harmonic mean. $F1 - score = \frac{2*(Precision*Recall)}{(Precision+Recall)}$

B. Experimental Setup

The data augmentation process is applied to increase the number of images in the dataset to ensure the algorithms have enough samples for training. Hence, the rotated, zoomed, and horizontally flipped images have been added to the dataset before training. Experiments involve four bird species datasets (CUB-200-2011, Kaggle 510 Bird Species Dataset, 325 Bird Species Dataset, and 100-Malaysia-Birds Dataset).

Several parameters must be set before fitting the dataset in the training process. The optimizer of the model is Stochastic Gradient Descent (SGD), a popular optimizer in the deep learning model. It iteratively changes the model's parameters depending on the gradient of the loss function concerning the parameters by using small batches of training samples. Although it may necessitate the careful selection of the learning rate and can profit from alternative methods for enhanced convergence and stability, SGD is computationally efficient and ideal for large datasets. Hence, the learning rate is 0.0001, and the momentum is 0.9.

In the Inception-V3 algorithm, there is a fine-tuned process that freezes the first 249 layers in the model and sets them with low epoch numbers which are 3, to make the algorithm adapt to the features in the dataset. After finishing with 3 epochs, all of the layers from the model are unfrozen and trained with the larger epochs, 10, to train more accurately. For the EfficientNet-B4 model, there is also a fine-tuned process that freezes the first 469 layers and trains with 3 epochs. After that, the remaining layers are unfreezing and retrained with new features obtained in 10 epochs.

The early stop is also included in both deep learning algorithms to prevent overfitting and monitor validation loss. The training progress will be stopped if the validation loss is not decreased. All the training process is performed using the Kaggle notebook under the GPU-100. The models used are implemented using Keras, and the batch size is set to 32, image size is 224x224.

C. Results and Analysis

Performance of Inception-V3 model for four different bird species dataset is displayed in Table III. Based on Table III, the best performance is obtained using 325 Birds Species Dataset, followed by the Kaggle 510 Bird Species Dataset and 100-Malaysia-Birds Dataset. These three datasets have achieved more than 90% accuracy as they have enough training samples for the model. CUB-200-2011 gets the poorest accuracy due to the small training samples, where it only consists of a maximum of 60 images per species.

TABLE III
THE PERFORMANCE OF THE INCEPTION-V3 MODEL

Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Time per epochs(s)
CUB-200-2011 Dataset	69	70	69	69	110
Kaggle 510 Bird Species Dataset	96	97	96	96	664
325 Bird Species Dataset	98	98	98	98	224
100-Malaysia-Birds Dataset	91	92	91	91	234

Meanwhile, in Table III, the execution time increases when the dataset's size increases. For instance, the time per epoch in the Kaggle 510 Bird Species Dataset is 664s, and it takes

1.8 hours to finish the training. Besides, the 325 Birds Species Dataset and 100-Malaysia-Birds Dataset have the same execution time in each of the epochs. However, the execution time of the CUB-200-2011 is the least as it only provides 40 images for the model training in each epoch.

TABLE IV
THE PERFORMANCE OF THE EFFICIENTNET-B4 MODEL

Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Time per epochs(s)
CUB-200-2011 Dataset	74	74	74	73	134
Kaggle 510 Bird Species Dataset	99	99	99	99	1229
325 Bird Species Dataset	99	99	99	99	719
100-Malaysia-Birds Dataset	92	92	92	92	265

Table IV presents the performance of the EfficientNet-B4 model on four distinct bird species datasets. The accuracy achieved by the EfficientNet-B4 model on the Kaggle 510 Bird Species Dataset and the 325 Birds Species Dataset is 99% across four evaluation metrics. This high accuracy can be attributed to the larger model size, which allows for more layers and enhanced training capabilities. Furthermore, the results demonstrate that larger datasets lead to optimal performance in deep learning models. Similarly, the Inception-V3 model shows a similar trend, with the CUB-200-2011 dataset exhibiting poorer performance but a 5% increase in accuracy when using a larger model. Moreover, employing a larger model also improves the overall performance of the 100-Malaysia-Birds Dataset.

The execution time for all datasets increases due to the larger network model (EfficientNet-B4) in Table IV compared to Inception-V3 in Table III. The Kaggle 510 Birds Species Dataset exhibits the longest execution time, with each epoch taking 1229 seconds and 34 hours for 10 epochs. This dataset's larger number of images contributes to the longer execution time. Similarly, the second largest dataset, the 325 Birds Species Dataset, requires 719 seconds per epoch for training, which is a substantial increase compared to the Inception-V3 model. In summary, the larger neural network model requires a longer training time than the smaller deep learning model.

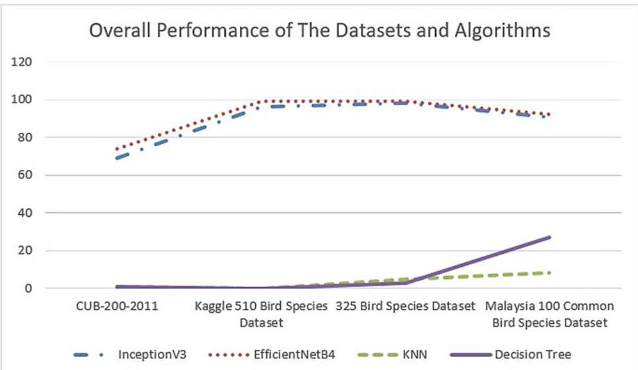

Fig. 4 Overall performance of algorithms in the experiment

Fig. 4 illustrates the comprehensive performance of four datasets in the experiment, comparing accuracy scores between machine learning techniques (K-Nearest Neighbor and Decision Tree) and deep learning approaches (Inception-V3 and EfficientNet-B4). The reason for performing this comparison is to prove that deep learning is better than the traditional machine learning approaches. The results in Fig. 4 indicate that deep learning approaches perform excellently, whereas machine learning techniques necessitate extensive pre-processing techniques [33].

D. *Comparison Results*

The comparison of the proposed method with other models is depicted in Table V. In Table V, Ragib et al. [21] used the pre-trained deep learning models, ResNet-18 and ResNet-101 for bird species recognition and achieved 96.71% and 97.98% accuracy for both methods, respectively. However, they only used the dataset's 15 out of 200 species to calculate the performance. This made the results not accurate. Alswaitti et al. [22] used the pre-trained network, AlexNet, DenseNet-121, GoogLeNet, and ResNet-50, and obtained the results up to 98% accuracy. They only used 180 bird species and 20000 images from the dataset, where the model might not recognize most birds. Furthermore, Jha et al. [23] received 84.76% accuracy score using sequential CNN to perform the bird recognition. This author did not mention how many classes were used in the proposed bird species recognition system.

TABLE V
COMPARISON WITH PREVIOUS WORKS

	Model	Accuracy (%)
Ragib et al. [21]	ResNet-18	96.71
	ResNet-101	97.98
Alswaitti et al. [22]	AlexNet	93.40
	DenseNet-121	98.60
	ResNet-50	97.70
Jha et al. [23]	Sequential CNN	84.76
The proposed method	InceptionV3	98
	EfficientNetB4	99

In Table V, the proposed method demonstrated outstanding performance compared to previous works. While Ragib et al. [21] and Alswaitti et al. [22] also employed pre-trained deep learning models, they did not undergo fine-tuning, potentially leading to imperfect feature extraction from the dataset. The poorest result was obtained by Jha et al. [23], who solely utilized a sequential CNN model for training. This model begins with random weights, and the training process focuses on the convolutional layers the users set. The drawback of this model is that there is no ultimate answer on how many layers need to be set, and it depends on the dataset's quality. In summary, fine-tuned deep learning models have the potential to achieve superior performance compared to both pre-trained models without fine-tuning and sequential CNN models.

E. *Bird-predicted System with Graphical User Interface (GUI)*

Upon completing the training process, the model exhibiting the highest accuracy (EfficientNet-B4) is chosen to be integrated into the bird-predicted system with Graphical User Interface (GUI), enabling users to predict the bird species based on input images. The system can pre-process the input image and pass it through the model for prediction. The system then presents the predicted bird species' names and the corresponding confidence score. Furthermore, the system can predict the bird species' name even when the uploaded bird image is rotated or blurred format. The system output and testing results are depicted in Fig. 5 through Fig. 7.

Fig. 5 Bird-predicted system with Graphical User Interface (GUI)

Fig. 6 Bird-predicted system using rotated image

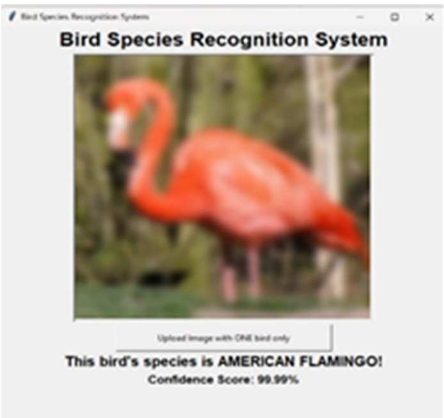

Fig. 7 Bird-predicted system using blur image

IV. CONCLUSION

This paper has employed two fine-tuned deep learning algorithms, demonstrating superior performance in large datasets. An adequate number of images per label is crucial for the model to learn the features in the images effectively. The more features the model learns, the higher accuracy it achieves. For instance, the Kaggle 510 Birds Species Dataset and the 325 Birds Species Dataset achieve over 90% accuracy thanks to their extensive collection of over 50,000 images evenly distributed across bird species. Besides, the size of the deep learning model is another factor influencing performance in this task. Increasing the number of layers in the network allows more features to be learned at different levels. Moreover, the developed GUI facilitates the prediction of various bird species, including rare ones not commonly encountered in daily life, and provides users with confidence scores for each predicted image.

ACKNOWLEDGMENT

The research is supported by Multimedia University Internal Fund through the IR Fund 2021 (MMUI/210029).

REFERENCES

[1] C. Collins, D. Dennehy, K. Conboy, P. Mikalef, “Artificial intelligence in information systems research: A systematic literature review and research agenda,” *International Journal of Information Management*, vol. 60, 2021, doi: 10.1016/j.ijinfomgt.2021.102383.

[2] MC. Tai MC, “The impact of artificial intelligence on human society and bioethics,” *Tzu Chi Medical Journal*, vol. 32, 2020, pp. 339-343, doi: 10.4103/tcmj.tcmj_71_20.

[3] P. Bhatt and A. Muduli, “Artificial intelligence in learning and development: a systematic literature review,” *European Journal of Training and Development*, vol. 47, 2023, pp. 677-694, doi: 10.1108/EJTD-09-2021-0143.

[4] Y. M. Rosas, P. L. Peri, J. Benítez, M. V. Lencinas, N. Politi, and G. M. Pastur, “Potential biodiversity map of bird species (Passeriformes): Analyses of ecological niche, environmental characterization and identification of priority conservation areas in southern Patagonia,” *Journal for Nature Conservation*, vol. 73, 126413, 2023, doi: 10.1016/j.jnc.2023.126413.

[5] L. Cardador and T. M. Blackburn, “A global assessment of human influence on niche shifts and risk predictions of bird invasions,” *Global Ecology and Biogeography*, vol 29, no. 11, 2020, pp. 1956–1966, doi: 10.1111/geb.13166.

[6] B. Naimi, C. Capinha, J. Ribeiro, C. Rahbek, D. Strubbe, L. Reino, M. B. Araújo, "Potential for invasion of traded birds under climate and land-cover change," *Global Change Biology*, pp. 1-13, 2022, doi: 10.1111/gcb.16310.

[7] C. H. Oliveros, D. J. Field, D. T. Ksepka, et al., “Earth history and the passerine superradiation,” in *Proceedings of the National Academy of Sciences of the United States of America*, vol 116, no. 16, pp. 7916-7925, 2019, doi: 10.1073/pnas.1813206116.

[8] A. E. Barnes, J. G. Davies, B. Martay, P. H. Boersch-Supan, S. J. Harris, D. G. Noble, J. W. Pearce-Higgins and R. A. Robinson, “Rare and declining bird species benefit most from designating protected areas for conservation in the UK,” *Natural Ecology & Evolution*, vol. 7, pp. 92–101, 2023, doi: 10.1038/s41559-022-01927-4.

[9] X. Yang, S. Li, A. Hughes and G. Feng, “Threatened bird species are concentrated in regions with less historical human impacts,” *Biological Conservation*, vol. 255, 2021, doi: 10.1016/j.biocon.2021.108978.

[10] M. M. Taye, "Theoretical Understanding of Convolutional Neural Network: Concepts, Architectures, Applications, Future Directions," *Computation,* vol 11, no. 3:52, 2023, doi: 10.3390/computation11030052.

[11] J. Florentin, T. Dutoit, and O. Verlinden, “Detection and identification of European woodpeckers with deep convolutional neural networks,” *Ecological Informatics*, vol. 55, p.101023, 2020, doi: 10.1016/j.ecoinf.2019.101023.

[12] A. Khot, "Image analysis using convolutional neural network to detect bird species," *7th International Conference on Computing in Engineering & Technology (ICCET 2022)*, Online Conference, 2022, pp. 58-61, doi: 10.1049/icp.2022.0592.

[13] L. Alzubaidi, J. Zhang, A.J. Humaidi, et al., “Review of deep learning: concepts, CNN architectures, challenges, applications, future

directions," *Journal of Big Data*, vol 8, no. 53, 2021, doi: 10.1186/s40537-021-00444-8.

[14] I. H. Sarker, "Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions," *SN Computer Science*, vol. 2, p. 420, 2021, doi: 10.1007/s42979-021-00815-1.

[15] B. Qiao, Z. Zhou, H. Yang and J. Cao, "Bird species recognition based on SVM classifier and decision tree," in *2017 First International Conference on Electronics Instrumentation & Information Systems (EIIS)*, Harbin, China, 2017, pp. 1–4, doi: 10.1109/EIIS.2017.8298548.

[16] P. Gavali and J. S. Banu, "Bird species identification using deep learning on GPU platform," in *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*, Vellore, India, 2020, pp. 1–6, doi: 10.1109/ic-ETITE47903.2020.85.

[17] N. Jain, M. Kamble, A. Kanojiya and C. Jage, "Implementation of bird species detection algorithm using deep learning," in *ITM Web Conference*, vol. 44, 2022, pp. 1–6, Art. no. 03042, doi: 10.1051/itmconf/20224403042.

[18] N. Sharma, A. Vijayeendra, V. Gopakumar, P. Patni and A. Bhat, "Automatic identification of bird species using audio/video processing," in *2022 International Conference for Advancement in Technology (ICONAT)*, 2022, pp. 1–6, doi: 10.1109/ICONAT53423.2022.9725906.

[19] H. Wang, Y. Xu, Y. Yu, Y. Lin, and J. Ran, "An efficient model for a vast number of bird species identification based on acoustic features," *Animals*, vol. 12, no. 18, p. 2434, 2022, doi: 10.3390/ani12182434.

[20] R. Pahuja and A. Kumar, "Sound-spectrogram based automatic bird species recognition using MLP classifier," *Applied Acoustics*, vol. 180, 2021, Art. no. 108077, doi: 10.1016/j.apacoust.2021.108077.

[21] K. M. Ragib, R. T. Shithi, S. A. Haq, M. Hasan, K. M. Sakib and T. Farah, "PakhiChini: Automatic bird species identification using deep learning," in *2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4)*, London, UK, 2020, pp. 1–6, doi: 10.1109/WorldS450073.2020.9210259.

[22] M. Alswaitti, L. Zihao, W. Alomoush, A. Alrosan and K. Alissa, "Effective classification of birds' species based on transfer learning," *International Journal of Electrical and Computer Engineering (IJECE)*, no. 4, vol. 12, 2022, pp. 4172–4184, doi: 10.11591/ijece.v12i4.pp4172-4184.

[23] D. Jha, and M. Sundaram, "Bird species identification using deep learning and image processing," *EasyChair*, 2022, Preprint no. 8319.

[24] J. Gómez-Gómez, E. Vidaña-Vila and X. Sevillano, "Western Mediterranean Wetland Birds dataset: A new annotated dataset for acoustic bird species classification", *Ecological Informatics*, vol. 75, 2023, Art. no. 102014, doi: 10.1016/j.ecoinf.2023.102014.

[25] H. K. Kondaveeti, P. Nithiyasri, B. S. L. Sri, K. H. Jessica, S. V. S. Kumar and S. C. Gopi, "Bird Species Recognition using Deep Learning," *2023 3rd International conference on Artificial Intelligence and Signal Processing (AISP)*, VIJAYAWADA, India, 2023, pp. 1-6, doi: 10.1109/AISP57993.2023.10134804.

[26] S. V. S. Kumar and H. K. Kondaveerti, "A Comparative Study on Deep Learning Techniques for Bird Species Recognition," *2023 3rd International Conference on Intelligent Communication and Computational Techniques (ICCT)*, Jaipur, India, 2023, pp. 1-6, doi: 10.1109/ICCT56969.2023.10075901.

[27] Y. -P. Huang and H. Basanta, "Bird Image Retrieval and Recognition Using a Deep Learning Platform," in *IEEE Access*, vol. 7, pp. 66980-66989, 2019, doi: 10.1109/ACCESS.2019.2918274.

[28] MY, Xin, L. W. Ang, and S. Palaniappan, "A Multi-Scale Feature Attention Image Recognition Algorithm," *Journal of Informatics and Web Engineering* (JIWE), vol. 2, no. 2, pp. 1-7, 2023, doi: 10.33093/jiwe.2023.2.2.1.

[29] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," *Computer Vision and Pattern Recognition*, 2015, pp. 2818-2826, doi: 10.48550/arXiv.1512.00567.

[30] A. F. Agarap, "Deep Learning using Rectified Linear Units (ReLU)," *Neural and Evolutionary Computing (cs.NE),* 2019, doi: 10.48550/arXiv.1803.08375.

[31] M. Wang, S. Lu, D. Zhu, J. Lin and Z. Wang, "A High-Speed and Low-Complexity Architecture for Softmax Function in Deep Learning," *2018 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS)*, Chengdu, China, 2018, pp. 223-226, doi: 10.1109/APCCAS.2018.8605654.

[32] M. Tan, and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks, " in *Proceedings of the 36th International Conference on Machine Learning*, Long Beach, CA, USA, 2019, pp. 6105-6114, doi: 10.48550/arXiv.1905.11946.

[33] C-C. Wong, L-Y. Chong, S-C. Chong, and C-Y. Law, "QR Food Ordering System with Data Analytics", *Journal of Informatics and Web Engineering (JIWE)*, vol. 2, no. 2, pp. 249–272, 2023, doi: 10.33093/jiwe.2023.2.2.18.