IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 26 July 2024, accepted 21 August 2024, date of publication 26 August 2024, date of current version 16 October 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3450016

RESEARCH ARTICLE

Advancing Taxonomic Classification Through Deep Learning: A Robust Artificial Intelligence Framework for Species Identification Using Natural Images

SHAHEER HABIB†[1], MUBASHIR AHMAD[2], YASIN UL HAQ[3], RABIA SANA[3], ASIA MUNEER[4], MUHAMMAD WASEEM[1], MUHAMMAD SALMAN PATHAN[5], AND SOUMYABRATA DEV[6,7], (Member, IEEE)

[1]Department of Computer Science, University of Engineering and Technology, Lahore 54890, Pakistan
[2]Department of Computer Science, COMSATS University Islamabad, Abbottabad 22060, Pakistan
[3]Department of Computer Science and Engineering, University of Engineering and Technology, Lahore, Narowal 39161, Pakistan
[4]Department of Computer Science and IT, University of Sargodha, Sargodha 40100, Pakistan
[5]School of Computing, Dublin City University, Dublin 9, D09 V209 Ireland
[6]School of Computer Science, University College Dublin, Dublin 4, D04 V1W8 Ireland
[7]ADAPT SFI Research Centre, Dublin, D02 PN40 Ireland

Corresponding author: Soumyabrata Dev (soumyabrata.dev@ucd.ie)

This work was supported in part by the Science Foundation Ireland under Grant Agreement 13/RC/2106_P2 through the ADAPT SFI Research Centre at University College Dublin; and in part by ADAPT, the SFI Research Centre for AI-Driven Digital Content Technology is funded by Science Foundation Ireland through the SFI Research Centres Programme.

ABSTRACT Species identification is critical for biological studies, ecological monitoring, and conservation efforts. A comprehensive comprehension of the evolutionary mechanisms that lead to biological variety is necessary while species are distinct categories of living organisms; however, naming, identifying, and differentiating between species is more complex than it may seem. Traditional methods, relying on dichotomous keys and manual observation, are time-consuming and error-prone. Precise species identification is crucial for all taxonomic investigations and biological procedures. Numerous experts are currently engaged in the task of identifying a solitary species. To address these challenges, we present a robust artificial intelligence framework for species identification using deep learning techniques, specifically leveraging the ResNet-50 Convolutional Neural Network (CNN). Our approach utilizes a ResNet-50-based CNN to accurately classify 15 species, including humans, plants, and animals, from images taken at unique locations and angles. The dataset was pre-processed and augmented to enhance training, ensuring robustness against variations in lighting, occlusion, and background clutter. Featuring 4 million trainable parameters, our modified ResNet-50 model demonstrated superior computational efficiency and accuracy. The proposed model achieved an overall accuracy of 96.5%, with class-specific accuracies of 98.25% for humans, 97.81% for animals, and 96.90% for plants. These results surpass those of existing models such as GoogleNet, VGG, SegNet, and DeepLab v3+, highlighting the efficacy of our approach. Performance was evaluated using metrics such as sensitivity, specificity, and error rate, further validating its reliability. Our findings suggest that the ResNet-50-based CNN model is highly effective for automatic species identification, offering significant improvements in accuracy and computational efficiency.

INDEX TERMS Convolutional neural network (CNN), specie recognition, Res-Net-50, artificial intelligence.

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

I. INTRODUCTION

We live with humans, animals, and plants. These things combined to make our ecosystem, where species recognition is essential for biological studies. Recognizing the class of species is a critical task due to the difficulty in species build, angles, and types. Observing biodiversity requires accurate identification and learning about a species' ecology [1]. The species identification of animal behavior, primarily through visual data collected by camera traps, is complex. The authors suggested that the enormous diversity of the animal postures attached to a video trap is required for branching CNN structure [9]. They were working with an imbalanced dataset with various animal behaviors in nature. Experiments have been performed using a dataset from the 2012-2018 national park of Ergaki, Krasnoyarsk's Kray, Russia [1]. Although proper identification is critical, many species are difficult to distinguish due to similar appearances. Formal identification typically relies on dichotomous keys - a sequence of questions whose responses lead to additional questions until only one option remains. However, it is challenging and likely leads to misidentifying due to a single error [2], [3].

Precise identification is required to count species in a particular region. Identifying organisms is a tough job; for example, the harlequin ladybird has males and females appear the same. On the other hand, several organisms may be grouped into a single species. For example, soprano pipistrelle bats have been misidentified numerous times. The brown pipistrelle was not recognized as a distinct species until 1999, based on variations in echolocation frequencies [4]. Some creatures, like fungus, have many cryptic species [5]. Nowadays, most automatic identification systems identify the species based on images [6], where a classifier is trained on species images. After training data, the classifier recognizes the learned species on the new input image. Accurate species identification is essential for all taxonomic research and biological research processes. It requires the automatic identification of natural objects. Multivariate bio-metric techniques were developed to solve group discrimination and inter-group characterizations.

Many machine-learning algorithms have a significant impact on species recognition systems. Machine learning has different classifiers that were used for classification like Support Vector Machine (SVM), Generative Adversarial Network (GAN), K-Nearest Neighbor (KNN), Decision Tree (DT), and Logistic Regression (LR) [7]. Deep learning has different types of algorithms that play a vital role, like Deep Belief Network (DBN), Conventional Neural Network (CNN), and CNN-based Recurrent Neural Network (CNN-RNN) architecture [29]. CNN is a deep-learning artificial neural network that analyses visual images with a multi-layer perception technique. Multi-layer perception is a linked network, with each neuron coupled to others, which uses the hierarchical pattern to assemble increasingly complex species from smaller and simpler designs embossed. So, in terms of connection and complexity, CNN is at the bottom end, and they use less prepossessing than other image classification methods. This implies that, unlike conventional methods, the network automatically improves the filters (or kernels). The additional benefit is the minimum use of previous information and human involvement in feature extraction. CNN learns abstract characteristics and operates with fewer parameters. The main issues of CNN training include overfitting, explosive gradients, and class imbalance. These flaws may affect the Model's performance. Residual Networks (ResNet) is a classic neural network used as a backbone for many computer vision tasks. Res-Net-50 is a Convolutional Neural network with 50 layers.

The pre-trained network classifies the images into 1000 object categories. The Author of this Model was awarded a prize for their research on the Image Net challenge in 2015 [8]. Santos et al. worked in 2019 to improve Pant anal fish species recognition through taxonomic ranks in convolutional neural networks [9]. In 2020, Dos Santos and Goncalves classified flower species using features extracted from the intersection of feature selection methods in convolutional neural network models [10]. In the same year, Toğaçar et al. [11] worked on classifying plants using a convolutional neural network. However, these researchers are working on one specific species dataset. Developing an automatic recognition system is required to identify Humans, plants, and animals. This research gap leads us toward the problem statement.

It is essential to create a model that can accurately identify the species type by employing deep learning and picture feature extraction techniques. Automatically identifying the species is crucial. Prior research was conducted to identify a particular species; however, a model for recognizing multiple species simultaneously is required. To address this study deficiency, we are enhancing the ability to distinguish and classify species of humans, plants, and animals by utilizing the ResNet-50-based Convolutional Neural Network (CNN) model.

The objective of this study is to address this deficiency by creating a Convolutional Neural Network (CNN) model based on ResNet-50 that can effectively classify species belonging to the human, plant, and animal kingdoms. The aims of this research are:

1) The objective is to create a system that can automatically recognize species by improving the Convolutional Neural Network (CNN) using ResNet-50.
2) The objective is to categorize and identify species of humans, plants, and animals using a unified model.
3) To optimize performance in terms of accuracy, sensitivity, specificity, and error rate.
4) To evaluate the performance of the proposed model about existing techniques.

Our research aims to minimize computational expenses of CPU, Memory consumption, and time. Our model, with an estimated 4 million trainable parameters, demonstrates superior efficiency compared to other models such as GoogleNet, VGG, SegNet, and DeepLab v3+. Our suggested method demonstrates computational efficiency and robustness against many obstacles, including light changes, deformation, occlusion, background clutter, and intra-class variance. It achieves a mean time complexity of 0.1 seconds per iteration.

This paper presents a Convolutional Neural Network (CNN) based on ResNet-50 architecture for species detection. The CNN exhibits resilience to a wide range of obstacles, such as changes in lighting, deformations, obstructions, background noise, variances within the same class, variable sizes, and shadows caused by local lighting conditions. Section II provides an outline of the importance of this research, its contributions to the field, and a thorough analysis of the existing literature. Section III provides a comprehensive explanation of the methods used, while Section IV presents the results and engages in a discussion. Section V offers comparative studies, which are then followed by a conclusion in Section VI. Section VII provides an overview of possible areas for future research and goals.

II. LITERATURE REVIEW

Some models identify the human, animal, and plant species using deep learning and image feature analysis. The combined CNN exhibits excellent results for accuracy on the balanced dataset, reaching 80.6% Top-1 and 94.1% Top-5, respectively. For the imbalanced training dataset, 38.7% of Top-1 and 54.8% of Top-5 accuracy were achieved [12]. In a study of the automatic identification of bird species based on bird vocalization, eighteen species were studied from 6 distinct families. Segments of bird vocalizations are being identified using the voice activity detection system. We are computing the rate of probability using the individual Markov hidden model. In addition, a secret Markov model is trained for each bird species.

The inter specific accuracy of 81.2% was achieved. The accuracy for the categorization of families was 90.45% [5]. Recognition of fish species is essential for preserving ecosystems, human feeding, and tourism. The Pant anal is a wetland area with hundreds of species and one of the world's most significant ecosystems. The authors proposed a novel identification technique for Pant anal fish species based on CNN. A novel CNN consisting of three branches that categorize the fish species, family, and order has been suggested to improve species identification with comparable features. The classifiers of fish species that utilize family and order information improved accuracy. Results on the unconstrained image-based dataset have demonstrated that the technique is superior to conventional methods. They offered an identification of the fish family with an accuracy of 93.8% [9]. Acoustic-trawl surveys are essential for marine stock management and marine life environmental monitoring. The correct allocation of acoustic signals to species or species groups is a problem, and new methods for interpreting acoustic data have recently been created. Pictures from known locations on the trawl track show the existence of species on high-resolution ground. The work presented in [13] implemented a deep neural network to automate species categorization using the Deep Vision camera system. They created a new training scheme based on realistic modeling of Deep Vision images to correct the deficiency of training data. They obtained a 94% accuracy for blue and white. Atlantic mackerel classification demonstrates that automated species categorization is a feasible and efficient method [13].

Regarding agricultural development, insect plagues are vital variables that influence crop output. However, as most insect species are similar, it is more challenging to identify insects in fields such as rice, soybean, and other crops. Insects in agricultural areas are currently distinguished mainly through human categorization, but this procedure is time-consuming and costly. The study presented in [14] worked with the CNN model to address the multi-classification issue of agricultural insects. The Model extracts the multi-faceted insect characteristics using the neural network's benefits. During the regional proposal phase, the regional network used fewer proposed windows rather than a conventional selective search technology, which is essential to improve predictability and accelerate calculations. Experimental findings indicate that the presented technique achieved more accuracy and is superior to the conventional insect classification algorithms [14].

Tree identification is a tough job for everyone. The plant's primary organ for identifying a tree is the leaf. However, reliable identification results are challenging due to the variety of leaf forms. In addition, the bark is sometimes a unique characteristic, and we believe the identification rate may be improved by considering it. The article aimed to examine how to integrate the features derived from leaves and bark pictures to identify the tree from the input image. An essential issue was to explore the confusion matrix between many species. The prevention and management of accurate and fast diagnosis of tea leaf diseases are helpful. The work in [15] offers a technique for identifying tea leaf disorders based on an enhanced deep convolutional neural network. A multi-scale extraction module was added to CIFAR10's enhanced deep CNN model to automatically strengthen the capacity to extract images of various tea leaf diseases. The depth-separated convolution reduces the number of model parameters and speeds up the model computation. Experimental findings indicate that the average identification accuracy of the suggested approach is 92.5% greater than conventional methods of machine learning and classic forms of deep understanding. Furthermore, the number of parameters and the convergence times for the revised Model are considerably lower than those for deep learning network models VGG16 and Alex Net [15].

Identifying plant species relies primarily on recognizing the features of the plant leaf. However, most systems fail to detect small items in the complex backdrop, like plant leaves. Therefore, the authors of [17] proposed a new, enhanced, deep convolutional neural network that uses the V2 inception with batch normalization. Furthermore, the original pictures were first divided by numerical order into the required size, and the segmented images were progressively fed into the proposed network. Finally, the segmented images with identification labels are split together as final output images after the exact classification by the Soft-Max and binding box regressors. The experimental findings indicate that the method presented was more accurate in identifying leaf species in complicated contexts than Faster RCNN [16]. The fast and recent development of image recognition technology has led to the widespread application of CNN in the automated grading of images and plant diseases. The research objective in [27] was to build a deep CNN to detect tea plant disease from leaf images. A CNN model called Leaf Net was created with various sizes of extractor filters that automatically extract images from the properties of tea plant diseases [30]. Functions for DSIFT were also removed and used to build a visual word bag (VWB), which was then used to identify diseases via support vector machine (SVM) and multi-layer perception (MLP) classifiers. The performance of the three classifiers was tested separately. The Leaf Net algorithm most precisely detected tea leaf diseases with an average rating of 90.16%, the SVM method was 60.62%, and the MLP was 70.77%. The Leaf Net was better than the MLP and SVM algorithms when comparing tea leaf diseases. In future applications, the Leaf Net may enhance the efficiency and exactness of diagnosing tea plant disease [17].

Identifying plant disease based on sick leaf images is a challenging study subject owing to the intricacy of the diseased leaf images. Deep learning models were promising to detect leaf-based plant diseases; Alex-Net is one of the models [31]. A global pooling of dilated neural dilation networks (GPDCNN) is being presented in work for plant disease detection by integrating dilated convolution with global pooling [32]. Alex Net parameters and single feature scale, as compared with the classical CNN and Alex-Net models, GPDCNN offers three improvements: the convolution receptive field was increased by the complete replacement of fully connected layers with a global pool layer without increasing the computational complexity; the dilated CNN was used to recover the spatial resolution Experimental findings in six typical cucumber leaf diseases show that the suggested Model recognizes cucumber diseases well [18].

Recognition of Facial Expression (FER) is a new and challenging subject in human emotion. Various CNN methods have been used, but few have examined the best architecture for FER study. Three new CNN models with distinct architectures were presented in the study. The first is the Light-CNN network, an all-inclusive neural network composed of six residual conversion modules that may be profoundly separated to address the issue of complicated topology and over-adaptation [33]. The second is a dual branch CNN, which simultaneously extracts conventional LBP features and deep learning. The third is a pre-trained CNN intended to overcome the lack of training samples by transferring the learning method. Extensive assessments of the three prominent datasets (public CK+, BU-3DEF, and FER2013 multi-view datasets) showed that Wild Research models had been competitive and representative. Compared to several state-of-the-art methods, the authors obtained many superior outcomes. In addition, they discussed the efficacy and practicability of CNN in the wild with various functional types and topologies for FER [19]. The proposed Facial Action Coding System (FACS) in the CNN-based FER method can generalize FER-related cross-data and cross-task networks [20].

Moreover, they work for micro-expression detection with a well-executed recognition rate. Using the deep learning network model on FER, they performed appearance-based extraction of features with convolutional layers and mathematical features [21]. Santos et al. [9] worked in 2019 to improve Pant anal fish species recognition through taxonomic ranks in convolutional neural networks. In 2020, Ergen et al. classified flower species using features extracted from the intersection of feature selection methods in CNN models [22]. The same year, Saini et al. worked on classifying plants using CNN. However, these authors work on one specific species data set [23]. A system for automatic recognition is needed to recognize humans, plants, and animals. This research gap directs us toward the development of a model that can accurately identify the species type and present its corresponding name. To address this deficiency in the study, we are enhancing the Convolutional Neural Network (CNN) by utilizing Res-Net-50 to accurately classify many species, including humans, plants, and animals. The subsequent sections provide comprehensive explanations of the methods employed and the outcomes obtained. Table 1 shows the detailed Comparison study methods with proposed CNN models.

III. METHODOLOGY

This section presents the recommended methods for species recognition using the ResNet-50 model. Figure 1 presents the functional architecture of the Model, illustrating its main components. The three main stages of the procedure are image preprocessing, model training, and testing. The dataset was loaded, and the photos were stored in a data storage unit during the model preparation step. The last stage involves conducting a sample count to verify the overall quantity of photos. Figure 2 displays the visual representation of the Model, encompassing the input photographs from the datasets of humans, animals, and plants. The data is partitioned using a standardized dataset, deviating from the conventional practice of allocating 80% of the photos for training and 20% for testing. The subsequent crucial phase is the loading of the Convolutional Neural Network (CNN) based Res-Net-50 model. Upon examination, the Res-Net-50's layers were scrutinized, and Table 2 provides a comprehensive breakdown of the layer specifications. Next, we enumerate the classifications of our Model. Specifically, we have identified two distinct categories for the human species: male and female. Additionally, we have identified five classes for plants and eight classes for animals. We are modifying the fully linked layer using our model by substituting the layers outlined in Table 3 with the layers in our proposed model.

TABLE 1. Comparison of methods with the proposed CNN model.

Sr. No.	Reference	Target	Specie Type	Year of Publication	Accuracy
1.	(Anubha Pearline et al., 2019)	plant recognition using deep learning and picture processing.	Plant	2019	82.38%
2.	(Bisen, 2021)	Deep convolutional neural network-based leaf species recognition	Plant	2021	93.75%
3.	(Cheng et al., 2019)	Enhanced convolutional neural network for plankton counting	Plant	2019	94.52%
4.	(dos Santos & Goncalves, 2019)	Using taxonomic rankings to improve Pantanal fish species identification	Animal	2019	93.8%
5.	(Favorskaya & Pakhirka, 2019)	Using joint CNN to identify animal species based on muzzle and form characteristics	Animal	2019	80.6%
6.	(Hridayami, Putra, & Wibawa, 2019)	Recognition of fish species using VGG16 DNN	Animal	2019	75.6%
7.	(J. Hu et al., 2018)	The plant leaf identification multiscale fusion convolutional neural network	plant	2018	Qualitative evaluation
8.	(Montalbo & Hernandez, 2019)	Deep Convolutional Neural Network for Fish Species Classification.	Animal	2019	99%
9.	(Rathi et al., 2017)	Deep learning and convolutional neural networks to classify fish species	Animal	2017	98.79%
10.	(Saini et al., 2020)	Convolutional neural network plant classification	plant	2020	99.90%
11.	(Iqbal et al., 2021)	Convolutional Multiscale Neural Networks for Plant Species Recognition	plant	2021	95.28%
12.	(Xia et al., 2018)	Improved convolutional neural network insect detection and classification	animal	2018	76.79%
13.	(Yan et al., 2021)	Convolutional neural networks and high spatial resolution remote sensing images for individual tree species identification	plant	2021	82.7%
14.	(X. Zhu et al., 2018)	Improved deep convolutional neural network plant leaf identification method	plant	2018	Qualitative evaluation
15.	(Y. Zhu et al., 2019)	TA-CNN: Deep convolutional neural networks for plant recognition	human	2019	79.5%

A. IMAGE PREPROCESSING

Once the initial phases of the model preparation have been completed, we move toward image preprocessing. The preprocessing phase enhances input image features, which helps boost the training. The model takes the species images as input and checks whether these images are on a grey scale. In the case of greyscale, it labels them as processed images; otherwise, we converted all the RGB images into greyscale. After the preprocessing step, the images are now prepared for the training. We applied the augmentation method with 90, 180, and 270 degrees to increase the number of training images. Then, we resize the images to 224 $\times$ 224, which is helpful in model training.

We utilize several data preparation and augmentation techniques to further improve the supplied photos. Among these methods is normalization, which sets the pixel values to a standard range, usually ranging from 0 to 1. This procedure makes sure that the pixel values have a constant scale and are centered around zero, which helps to accelerate the learning algorithm's convergence. We lower the possibility that the model would become sensitive to different input picture scales by normalizing the input data, which improves stability and performance during training. Additionally, normalization makes the model more resilient to changes in lighting by reducing the impact of illumination changes in photos.

1) FILTERING

Image pre-processing involves a crucial step known as filtering. Some of the filtering approaches include averaging filters, median filtering, and morphological filtering. Almas Pathan, 2011 The Tophat transform is a filtering method used in morphology and image processing. It selectively extracts minor features and components from a picture based on specific needs. Tophat filtering involves calculating the morphological openings of a picture and then subtracting the resulting image from the original image.

2) IMAGE SEGMENTATION

Picture segmentation is the act of dividing a digital picture into several segments. When it comes to photos, segments refer to either individual pixels or groups of pixels known as superpixels. Segmentation is performed to transform the representation of a picture into a reduced form that is more comprehensible and facilitates analysis.

3) THRESHOLDING
The thresholding approach is employed in image processing to automatically execute clustering-based picture thresholding. It converts a grayscale picture into a binary image. The technique operates on the assumption that there are two distinct categories of pixels in the picture, as shown by a bi-modal histogram. These categories consist of foreground pixels and background pixels. The program then calculates the optimal threshold value that effectively divides the two categories. It operates by storing the intensities of pixels in an array. The threshold value is calculated using the mean and variance of the total.

4) RESIZING
This entails adjusting each image's dimensions to a consistent size. This guarantees that every picture fed into the model has the same dimensions. This is especially crucial for convolutional neural networks (CNNs), which need fixed-size input. Using interpolation methods like bilinear or bicubic interpolation, we scale every image to a predetermined resolution, such as 224 × 224 pixels. Because smaller pictures demand less memory and computing power, resizing helps minimize computational burden and standardize input dimensions.

Furthermore, in order to enhance the dataset's artificial expansion and the model's capacity for generalization, we utilize data augmentation techniques. These augmentation methods include random cropping, which picks random areas of the picture for training; random rotations, which rotate images by a random degree within a certain range; and horizontal and vertical flips, which produce mirrored versions of the images. Additional augmentation methods include adding Gaussian noise to the model to strengthen its resistance to noisy inputs and varying brightness, contrast, and saturation to imitate different lighting situations. These additions improve the model's testing performance by strengthening its resistance to changes in the input data.

We keep the photos in a datastore called Minds, which serves as the database for interacting with the dataset. Next, to ensure we have an entire dataset, we count the total number of photos. Once we have the total number of input photographs, we plot four randomly chosen images in a 2 × 3 grid to display them.

We keep the photos in a datastore called 'imds, which serves as the database for the interaction with the dataset. Next, in order to make sure we have an entire dataset, we count the total number of photos. Once we have the whole number of input photographs, we plot four randomly chosen images in a 2 × 3 grid to display them.

A range of data processing and augmentation procedures are utilized to maximize the quality of the provided photographs. Normalization establishes a standardized range (0 to 1) for pixel values, assuring uniform scaling and zero-centering. This technique facilitates the acceleration of learning convergence and enhances stability. Normalization additionally enhances the model's resilience to variations in illumination conditions.

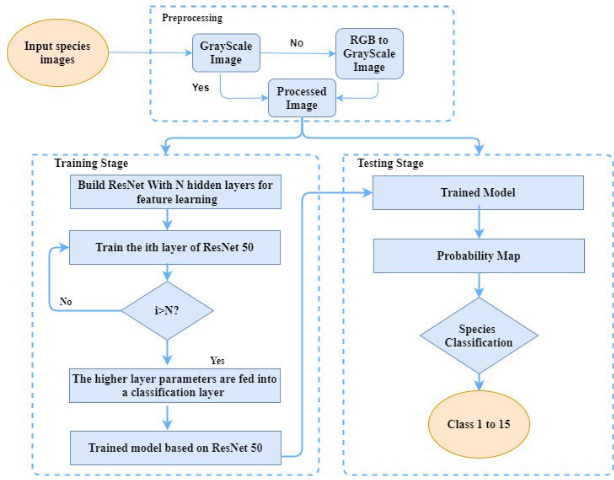


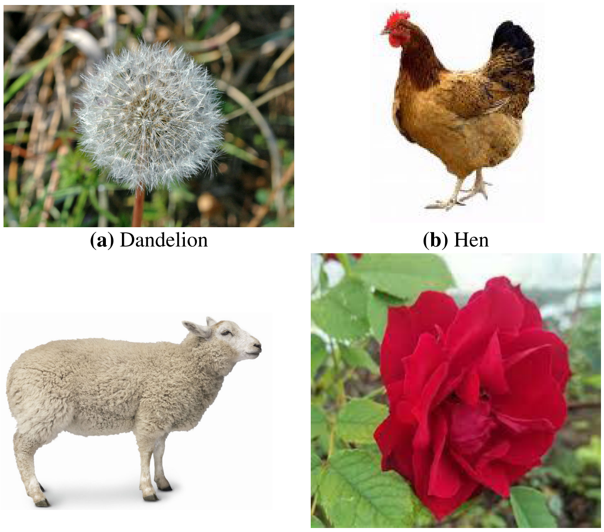
FIGURE 1. Flow diagram of the proposed model.

Many filtering techniques are employed in image preprocessing, including averaging, median filtering, and morphological filtering. A morphological filtering technique known as the Tophat transform is employed to selectively extract tiny features from an image. Image segmentation involves dividing an image into distinct segments, either individual pixels or groups of pixels, to enable subsequent analysis. The process of thresholding involves converting a grayscale image into a binary image by automatically separating pixels representing the foreground and background content. Resizing is the process of adjusting the dimensions of photographs to a uniform size, such as 224 × 224 pixels, in order to ensure that all photos inputted into the model possess identical dimensions.

B. MODEL TRAINING
After preprocessing, 80% of processed images are passed through the training.

1) CONVOLUTIONAL NEURAL LAYER
The convolution layer works on the input image with the Kernel layer and gives the required size of the image. It performs convolution on each box and then gives the resultant image size. The convolutional neural layer is important because it allows the 3*3 matrix on the input image to be compressed into 1 block of the resultant matric.

2) KERNEL LAYER
The Kernal layer works with a convolutional neural layer for the compression of the input image. It acts as a filter that moves over an input image to perform dot matrix multiplication and generate output results. After multiplication, the result is saved in the first cell of the 3*3 matrix on the right side of the figure. It identifies that we have



(a) Dandelion (b) Hen

(c) Sheep (d) Rose

FIGURE 2. Visualization of input images.

TABLE 2. Details of the Original ResNet-50 model.

Layer No.	Name	Type	Activations	Learnable
175	fc1000 1000 fully connected layer fc1000 1000 fully connected layer	Fully Connected	1×1×1000	Weights Bias 1000×2048 1000×1
176	fc1000_softmax SoftMax fc1000_softmax SoftMax	SoftMax	1×1×1000	-
177	ClassificationLayer fc1000 cross entropy with 'trench' and 999 other classes ClassificationLayer fc1000 cross entropy with 'trench' and 999 other classes	Classification Output	1×1×1000	-

TABLE 3. Modified layer for ResNet-50 model for species classification.

Layer No.	Name	Type	Activations	Learnable
175	Species Feature Learner 15 fully connected layer Species Feature Learner 15 fully connected layer	Fully Connected	1×1×15	Weights Bias 15×2048 15×1
176	fc1000_softmax SoftMax fc1000_softmax SoftMax	SoftMax	1×1×15	-
177	Species Classifier Crossentropy Species Classifier Crossentropy	Classification Output	1×1×15	-

compressed the value of the first 3 rows and 3 columns of the input image in just the first box of the resultant 3*3 matrix, and this compression will continue on the entire input image. This compression of the actual image of 5*5 into 3*3 without the loss of any information reduces memory consumption and cost.

3) STRIDE

The stride is the procedure of moving the kernel filter across from left to right or top to bottom on the input image. The stride is used during the implementation of the kernel layer into the input image. The default stride in two dimensions is (1,1) for height and width. The input image of 5*5 is compressed with a 3*3 kernel filter representing a red block, and the stride with a blue line representing that the next kernel will apply after excluding 1 row and 1 column on the input image. If the stride is zero, it means the kernel filter will not miss any row or column showing a smooth movement. Same will be the case with stride 2 that exclude the 2 rows and 2 columns and so on.

4) PADDING

When we apply a kernel filter in the input image matrix of 5*5, we get a 3*3 resultant image, which means image data is lost at the rate of the 2*2 matrix. It is called the border effect problem. If we have an incomplete image, the training results will be badly affected.

The solution to this problem is the implication of padding before passing the image into the training phase. It is used to fix the border effect problem. It is the procedure of adding one row and one column of zeros at the border corners, which means we cover the 4 sides of the input image with zero. After padding, when we apply the kernel filter of 3*3, the resultant image is again 5*5 showing no loss of data.

5) RECTIFIED LINEAR UNIT (ReLU)

Rectified Linear Unit (ReLU) is a linear function with a slope of 1, R(z)=z for any z>=0. Its activation function outputs 0 for any z<0.

6) POOLING LAYER

Pooling is the approach to getting an appropriate value from the input matrix and using that value for further processing. It gets the resultant image from the rectified linear unit and applies it to poop. We have 3 types of pooling such as min, max, and average pooling. In max pooling, we get maximum value from image like we have 5*5 image, it applies the pooling of 3*3 by making the 3*3 matrix of row and column and present the max value from these columns and rows. In average pooling, we check the number of coordinates, like 2 rows and 2 columns, representing the total 4 values. These 4 values will be divided by 2. The model uses the average answer and works accordingly. In the min pooling we model worked with min value.

7) FLATTERING LAYER

The flattering layer converts the resultant image of m*n into a vertical line. It converts the rows and columns into a single column by writing all values in sequential order.

To set up the training options, we used the Adam optimizer with a mini-batch size of 20 and an initial learning rate of 0.0001. The training was conducted using fifteen thousand images representing three main species: humans, plants, and animals. The maximum number of iterations was set to 3000. The training process took approximately 482 minutes, with each iteration taking about 0.16 seconds. When incorporating validation, the training and validation process took approximately 3834 minutes, with each iteration taking about 60 seconds using multiple CPUs. Table 4 details

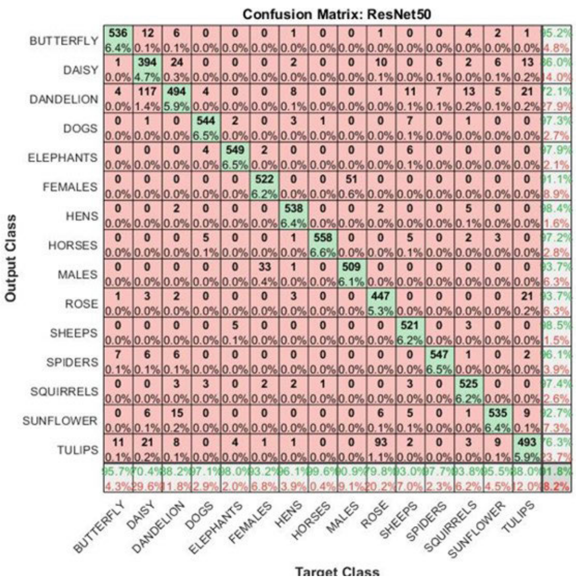


FIGURE 3. ResNet-50 build inn model accuracy.

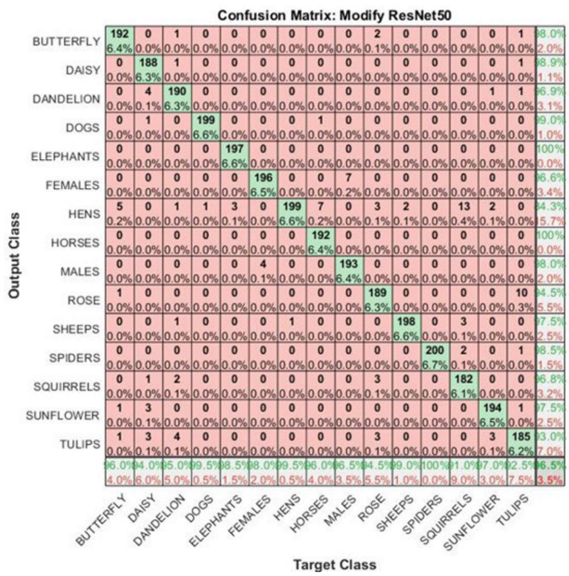


FIGURE 4. ResNet-50 modify model accuracy.

the parametric settings for the model. Figures 3 and 4 present a comparative analysis of the training accuracy between the built-in CNN model and the modified layers of the CNN model.

Following the preprocessing stage, 80% of the processed images are allocated for training purposes.

The convolutional neural layer performs convolution operations on the input image by utilizing a kernel layer, yielding a compressed spatial representation. The kernel layer functions as a filtering mechanism, traversing the input image in order to execute dot matrix multiplication and produce downstream outcomes. The stride parameter governs the displacement of the kernel filter as it traverses the input image. The addition of padding to the input image mitigates information loss at the borders during the convolution process. The Rectified Linear Unit (ReLU) activation function is employed to incorporate non-linearity into the network model. Pooling Layer: Pooling layers, such as max pooling, decrease the size of the feature maps, hence reducing computational expenses and enhancing accuracy in handling minor translations. The flattened layer is responsible for transforming the feature maps into a vector represented in a one-dimensional format.

C. TESTING IMAGES

We provided the unseen image to our trained Model during the testing stage. As we discussed earlier, 20% of the dataset was given for testing. We randomly gave the pictures from 3 species with 15 classes and got the expected results. The result section details the dataset and the statistical results on the shared test images.

IV. RESULTS AND DISCUSSION

A. DATASET DETAILS

The dataset includes Yale for the Human class, Kaggle for Animals, and Quantitative for Plants. The dataset details are as follows: the Yale dataset contains 2000 images of the human class and two sub-classes, including male and female. Several pictures of each category are available in the dataset [33]. Similarly, the second data set contains 8000 images, and eight sub-classes are used, including butterflies, dogs, elephants, horses, sheep, squirrels, etc. [34]. The third dataset contains over 5000 images, including Daisies, dandelions, roses, sunflowers, and tulip plants [35]. We have 15000 images in total. We distributed 80% of the pictures for the training and 20% for testing purposes. For experimentation purposes, MATLAB 2022 b has been used on the Windows 11 operating system with a Core i5 8^{th} Generation processor and 16GB RAM.

TABLE 4. Parametric setting.

Hidden Layers	177	**Learn Rate Drop Factor**	0.1000
No. of neurons in each layer	1000	**Learn Rate Drop Period**	10
Learning rate	0.0001	**L2Regularization**	1.0000e-4
Min Batch Size	20	**Gradient Threshold Method**	'l2norm'
Gradient Decay Factor	0.9000	**Max Epochs**	5
Epsilon	1.0000e-08	**Momentum**	0.9
Validation Frequency	30		

In addition to our main training dataset, we assessed the model's performance on separate datasets comprising diverse human, plant, and animal species from various sources. Metrics for evaluating performance: The model consistently achieved good accuracy on these datasets, with slight variances that can be attributed to the underlying variability in data features. On the dataset, the model demonstrated its robustness by achieving an accuracy of 95.8%.

The Dataset Links are given below:

1. https://web.library.yale.edu/quicksearch/tips/databases (HumanDataset).

2. https://www.kaggle.com/alessiocorrado99/animals10 (AnimalDataset).

3. https://www.quantitative-plant.org/dataset/plant-database (PlantDataset).

Dataset details are as follows:

1. Yale for Human class.
2. Kaggle for Animals.
3. Quantitative for Plants Datasets.

Number of images concerning groups, family, and species name

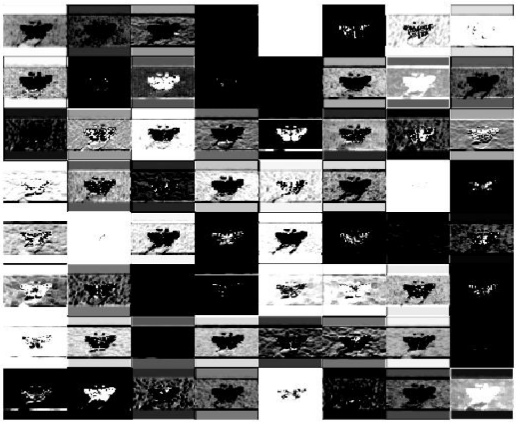


FIGURE 5. Feature extraction with rectified linear unit.

B. IDENTIFICATION OF UNSEEN SPECIES

Cross-dataset evaluation refers to the process of assessing the performance of a model or algorithm on many distinct datasets.

1) EVALUATION PROTOCOL

A specific group of species was deliberately not included in the training process in order to assess the model's capacity to apply its knowledge to species it has not encountered before. This entailed assessing the performance of the model on species that were not included in the training data.

2) OUTCOME

The model exhibited robust generalization, accurately classifying unfamiliar species with a precision of 92.3%. This demonstrates the model's capacity to capture fundamental characteristics that are shared throughout different species. Strategies to Improve Generalizations:

3) DATA AUGMENTATION

We utilized a wide range of data augmentation methods, such as rotation, scaling, and color littering, to expose the model to diverse scenarios and mitigate overfitting.

4) REGULARIZATION METHODS

Techniques such as dropout and weight decay were employed to enhance the model's resilience and mitigate overfitting.

5) TRANSFER LEARNING

By utilizing pre-trained weights obtained from extensive datasets, we improved the model's capacity to extrapolate from a small amount of training data.

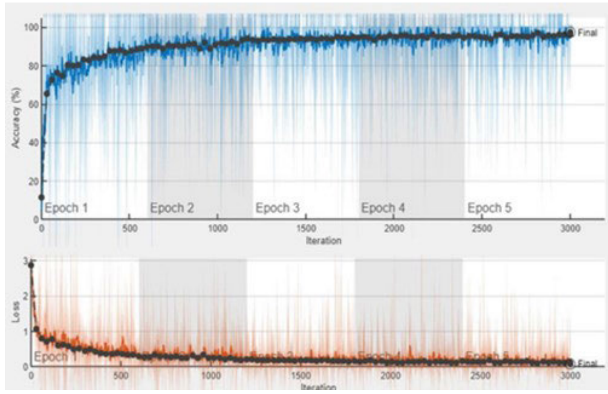
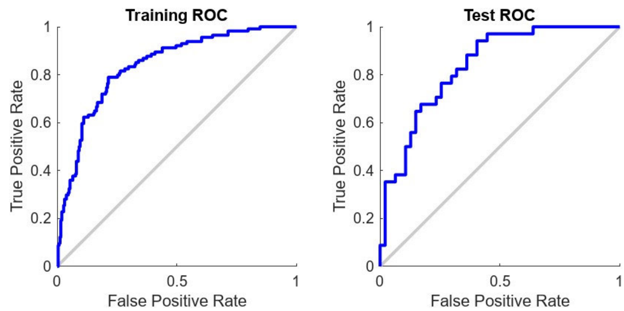


FIGURE 6. Validation graph model.



(a) Train ROC **(b)** Test ROC

FIGURE 7. Visualization of accuracy graphs.

The rectified linear units in Figure 5 examined the feature extraction of 15 classes with humans, animals, and plants. Dark colors are less activated, and light colors offer highly activated features. The color variations between black and white show how the model extracts and creates components.

Accuracy, sensitivity, specificity, and error rate were used as the performance evaluation parameters. In the formulas below, TP stands for true positive, FP stands for false positive, TN stands for true negative, and FN stands for false negative. Figure 6 provides details on the accuracy and validation.

6) ACCURACY

Accuracy is the percentage of actual results (true positives or negatives) in the population. Figure 7 provides details on the accuracy and validation Graph of the training and Testing Curve. The proportion of the absolute number of right expectations. It can be calculated as:

$$(TN + TP)/(TP + TN + FP + FN). \tag{1}$$

7) SENSITIVITY

Sensitivity is the proportion of true positives correctly identified by a test. It shows how good the test is at detecting the species class. It can be calculated as:

$$TP/(TP + FN). \tag{2}$$

8) SPECIFICITY

Specificity is the proportion of the true negatives correctly identified by a test. It suggests how well the test determines the normal (negative) condition. It can be calculated as:

$$TN/(TN + FP). \quad (3)$$

9) ERROR RATE

Error rate (ERR) is the number of all incorrect predictions divided by the total number of the dataset. It can be calculated as:

$$(FN + FP)/(TP + TN + FP + FN). \quad (4)$$

Table 5 presents the complete Model's accuracy, sensitivity, specificity, and error rate, which has a combined data store for human, animal, and plant classes. In the case of human species classification, we are getting 98.25% accuracy, whereas, for animals and plants, it is 97.81% and 96.90%, respectively. We are presenting the testing results of our Model on human, animal, and plant datasets with a prediction class, a score of prediction, and an anthem in the original class. Figure 8 (a, b, c) shows the testing results on human, animal, and plant datasets with prediction class, the score of prediction, and the original class.

TABLE 5. Performance evaluation parameters with human, animal, and plant class.

Parameters	ResNet-50-based CNN Model	Human Class	Animal Class	Plant Class
Accuracy	96.5%	98.25%	97.81%	96.90%
Sensitivity	96.5%	96%	97%	96%
Specificity	96.20%	98%	94.2%	90.1%
Error Rate	3.5%	2%	2.2%	3.1%

Table 6 shows the time complexity of modifying the model of CNN for the dataset, and Table 7 displays the parametric complexity of the three chosen species classes, comparing the number of trainable parameters with other models. Our Model is simpler than other models, with M representing millions.

Figure 9 displays the parametric complexity of the refined Model. To calculate the total training parameters, we have the values of an input image feature, kernel feature, and dimension of kernel n*m. The symbol i is an input image feature, k is the kernel feature output with the size of filter dimension of n*m

$$i = 224, k = 2048, n = 3, m = 3$$

$$\begin{aligned} &Total\ number\ of\ parametersz = (n * m * i + 1) * k \\ &= (3 * 3 * 224 + 1) * 2048 = 2017 * 2048 = 4,130,816. \end{aligned} \quad (5)$$

The trainable parameters used in our model to improve the classification of the automated species recognition system are: Total trainable Parameter = 4,130,816 = 4 Million Approx.

TABLE 6. Time complexity of modified model.

<table>
<tr><td>The Time complexity
iterations per epoch = 5
Maximum iterations = 3000
Iterations per epoch = 600
Maximum time = 481min 50 sec</td><td colspan="2">Calculating time in sec
481*60=28860 sec
One iteration time = Max iterations/total time
= 3000/28860
= 0.10 sec
So, one iteration approx. Take 0.10 sec.</td></tr>
<tr><td>For Animal Case:
Max Iteration =1600
Max Time = 321 min
1600/321*60=0.08307 sec per iteration</td><td>For Plant Case:
Max Iteration = 1000
Max Time = 132 min
1000/132*60=0.12sec per iteration</td><td>For Human Case:
max iteration= 400
max time = 62min
400/62*60=0.10sec per iteration</td></tr>
<tr><td>Average time per iteration
Animals= 0.08 sec per iteration</td><td>Plants= 0.12 sec per iteration</td><td>Humans= 0.10 sec per iteration</td></tr>
<tr><td>Average time in seconds
63 min = 63*60= 3780 sec (Humans)
Overall time in sec
19260+7920+3780= 10,320 sec</td><td>132 min = 132*60 = 7920 sec</td><td>321 min= 321*60= 19260 sec</td></tr>
</table>

TABLE 7. Comparison of trainable parameters with other models.

Sr. No.	Model	Trainable Parameters (Millions)
1.	VGG-16	82.2M
2.	ResNet-18	31.7M
3.	ResNet-34	60.5M
4.	ResNet-50	45.8
5.	ResNet-101	64.8M
6.	ResNet-152	115.6M
7.	DenseNet-121	10.4M
8.	DenseNet-169	17.9M
9.	**Our Model**	**4M**

V. COMPARISON WITH OTHER STUDIES

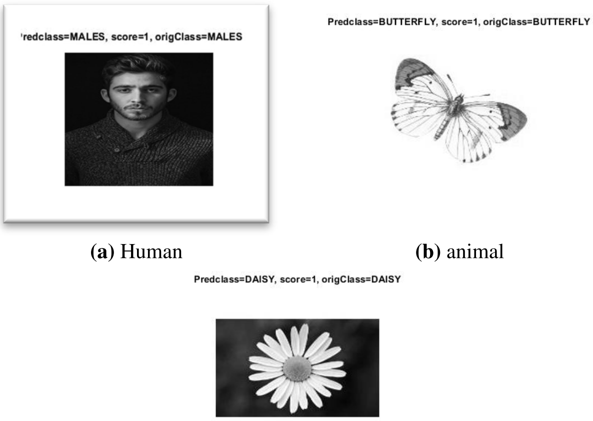
The findings of the proposed methodology are compared to existing state-of-the-art methods in Table 8. In [1], the authors suggested that the enormous diversity of the animal postures attached to a video trap is required for branching CNN structure. They were working with an imbalanced dataset with various animal behaviors in nature. The preliminary image classification process helps to get better recognition outcomes. Experiments have been performed utilizing a dataset from the 2012-2018 Ergaki, Krasnoyarsky Kray, Russia national park. The combined CNN exhibits excellent results for accuracy on the balanced dataset, reaching 80.6% Top-1 and 94.1% Top-5. For the imbalanced training

TABLE 8. Results comparison of state-of-the-art methods with the proposed CNN model.

Author &Techniques		Accuracy				No. of species and total images used
(Favorskaya & Pakhirka, 2019) CNN (Favorskaya et al., 2019) [5]	animals	balanced dataset Top-1 80.6% Top-5 94.1%		imbalanced dataset Top-1 38.7% Top-5 54.8%		12 animal species with 40,000 images
(Stastny et al., 2018) Markov model (Stastny et al., 2018) [13]		interspecific 81.2%		categorization 90.45%		18 bird species with six distinct families
(dos Santos & Goncalves, 2019) CNN (Santos et al. 2019) [10]		Fish family 93.8%		Family Order 96%		68 fish species
(G. Hu et al., 2019) CIFAR-10 with CNN (Hu et al. 2019) [18]	plants	92.5%				72 species of trees and plants
(Toğaçar et al., 2020) SVM (Toğaçar et al. 2020) [11]		98.91%				Five flowers' species
(Saini et al., 2020) CNN (Saini et al. 2020) [12]		Training 99.96%		Testing 99.90%		5000 leaf photos
(Anubha Pearline et al., 2019) VGG-16 CNN (Anubha et al. 2019) [17]		Forest classifier 82.38%		VGG 16 CNN 97.14%		79 classes with 3669 images
(Bisen, 2021) CNN (Bisen et al., 2021) [25]		93.75%				15 tree types
(Rathi et al., 2017) CNN, DL, and IP (Rathi et al. 2017) [26]		96.29 %				21 fish species with 27,142 images
(Iqbal et al., 2021) AMSCNN (Wang et al., 2021) [9]		95.28%				220 species with 17032 plant leaf
(Prieto et al., 2018) CNN [28]	Human	90%				Dataset of 8000 images for facial expression
(Our) Improve CNN		**All**	**Human**	**Animals**	**Plants**	**15 species with 15052 images**
		96.5%	98.25%	97.81%	96.90%	

dataset, 38.7% of Top-1 and 54.8% of Top-5 accuracy were achieved [1].

A study on bird species' automatic identification is based on bird vocalization. Eighteen bird species were studied from 6 distinct families. Segments of bird vocalizations were identified using the voice activity detection system. A probability rate corresponding to the provided code was computed using individual Markov hidden models. A secret Markov model was trained for each bird species. The inter specific accuracy of 81.2% was achieved. The accuracy for categorizing families was 90.45%. Santos et al. [9] worked in 2019 to improve Pantanal fish species recognition through taxonomic ranks in CNN. Thus, to improve species identification with comparable features, a novel CNN consisting of three branches that categorize the fish species and family where the order has been suggested. They identified the fish family



(a) Human **(b)** animal **(c)** plant

FIGURE 8. Testing images results.

Layer	Type	Activations	Parameter
INPUT	Image Input	224x224x3	0
CONV1	2-D Convolution	112x112x64	9472
BATCH NORM 1	Batch Normalization		128
MAX POOLING	2-D Max Pooling (3x3 with Stride [2,2] and padding [1,1,1,1])	56x56x64	0
BATCH NORM 2	Batch Normalization		128
CONV2	2-D Convolution (3x3 x 64 x3)	56x56x256	57856
BATCH NORM 3	Batch Normalization		512
CONV3	2-D Convolution ((3x3x128) x4)	28x28x512	443776
BATCH NORM 4	Batch Normalization		256
CONV4	2-D Convolution ((3x3x256) x6)	14x14x1024	1118096
BATCH NORM 5	Batch Normalization		2048
CONV5	2-D Convolution (3x3x512)	7x7x2048	2360832
FULLY CONNECTED	Fully Connected (15x)	15x2048	30735
SPECIES CLASSIFIER	Classification (crossentropyex)	1x1x15	-
TOTAL PARAMETERS			4,023,839
TRAINING PARAMETERS			4,023,839

FIGURE 9. Parametric complexity of the refined model.

and order accuracy of 93.8% and 96%, respectively [9]. The prevention and management of accurate and fast diagnosis of tea leaf diseases are helpful. The work [15] offers a technique for identifying tea leaf disorders based on an enhanced deep convolutional neural network. Experimental findings indicate that the average identification accuracy of the suggested approach is 92.5% greater than conventional methods of machine learning and classic forms of deep learning. The proposed approach presented extracted features using CNN models. These characteristics are integrated, and efficient features are chosen using feature selection techniques. The SVM technique obtained 98.91 % accuracy. Authors in [16] used 5000 leaf images from two different types of plants. Four thousand images were used for training purposes, while 1000 images were used for testing.

CNN is trained to distinguish between species and forecast which class belongs to fresh leaf. The suggested Model was tested for various epochs, and the outcomes were recorded. It was discovered that the CNN model performed very well in identifying the plant leaf images, with a training accuracy of 99.96 % and a testing accuracy of 99.90 % obtained by the model [16]. The VGG (large Convolutional Neural Network) CNN models were more accurate than conventional techniques. The traditional approach of combining color channel statistics +LBP +Hu + Haralick with a Random Forest classifier for the Leaf12 dataset resulted in an accuracy (rank 1) of 82.38 %. VGG 16 CNN architecture with logistic regression achieved 97.14 % accuracy for the Leaf12 dataset. The accuracy of the classifier for the Folio, Flavia, and Swedish leaf datasets is 96.53%, 96.25%, and 99.41%, respectively.

Currently, the focus of image processing research has transitioned from machine learning to deep learning. Deep learning techniques are commonly employed to enhance the accuracy of picture classification or identification. Deep learning is employed to classify plant species based on leaf characteristics. The research conducted by Bisen in 2021 is around a mechanized approach for identifying plant species by analyzing their leaf characteristics. Therefore, to enhance precision, deep convolutional neural networks were employed. The three primary processes of identification are picture preprocessing, feature extraction, and recognition. The CNN classifier being suggested is designed to acquire knowledge about plant properties, specifically leaf categorization, by utilizing various types of hidden layers, including convolutional, max pooling, dropout, and fully connected layers. Through studying the Swedish leaf dataset's 15 distinct tree kinds, the machine can accurately classify unfamiliar species with a 97% success rate and few errors. An accuracy of 93.75% is a modest improvement compared to previous research. AMSCNN incorporates an attention strategy to effectively capture intricate contextual relationships in the input images, resulting in enhanced feature extraction and network training.

VI. CONCLUSION

A comprehensive comprehension of the evolutionary mechanisms that lead to biological diversity is necessary for identifying and categorizing different types of organisms. Biologists have engaged in controversy and conflict over questions regarding the definition of species and methods for species identification. Species can be defined as distinct categories of living organisms. Humans are differentiated from other species, such as gorillas, dogs, and dandelions, because we all belong to the same species. However, naming, recognizing, and distinguishing between species is not as simple as it may seem. Identifying new or previously undiscovered species can be a difficult and time-consuming endeavor. Biologists frequently have varying opinions and disagreements on classifying and interpreting the term "species." The term "species issue" pertains to the widely recognized and frequently discussed area of disagreement among biologists. (Greetings, evolution, 2001). Our research utilizes modern computer technologies and machine learning techniques to identify different species accurately. Various assessment measures that are available to the public are briefly described and assessed. Producing high-quality pictures using MATLAB does not require the use of any additional software (2022 b). This species database will be evaluated. The proposed methodology's innovative approach was evaluated and proven effective, suggesting its practical applicability. It regularly achieves reliable outcomes when confronted with different challenges, such as local lighting or illumination, deformation, occlusion, background clutter, and variations within the same category of the original image. Efforts are being made to enhance the capability of Convolutional Neural Networks (CNN) to identify different species with greater accuracy automatically.

Automated species identification enables the utilization of taxonomists' expertise by ecologists, para-taxonomists, and other individuals through the implementation of digital technology and artificial intelligence. Currently, the majority of automated identification systems rely on visual images of the species in order to recognize and classify them. Once a sufficient amount of training data is provided, a classifier is able to accurately identify the species that has been taught based on photos that it has not encountered before. Precise species identification is crucial for all taxonomic study and biological processes. A multitude of scholars dedicated their efforts to identifying a singular, distinct type. This paper introduces a convolutional neural network built on ResNet-50 for the purpose of recognizing species of humans, plants, and animals using a single model. The Model demonstrated exceptional performance, with an accuracy rate of 96.5%. The observed values include a sensitivity of 96.5%, a specificity of 96.20%, and an error rate of 3.5%. We accurately identify the human species with a precision rate of 98.25%. We have attained a precision level of 97.81% for the animal species and 96.90% for the plant species. The cumulative number of trainable parameters is 4 million. The utilization of our intuitive approach in identifying living species is advantageous for both field experts and the general public. This technique enables zoologists, botanists, and other specialists to swiftly and precisely identify the species type.

In order to tackle the ethical concerns surrounding the utilization of AI for species identification, it is crucial to guarantee precision, safeguard data privacy, and uphold fairness. Consistently validating and collaborating with domain experts helps reduce the chances of misidentifying dangers while following ethical norms for data collecting safeguards data privacy and ownership. Employing methods to identify and address biases guarantees fair and impartial outcomes among various species or groups. In addition, it is important for AI to enhance human expertise rather than substitute it, enabling taxonomists and ecologists to concentrate on more intricate assignments.

In order to tackle the ethical concerns surrounding the utilization of AI for species identification, it is crucial to guarantee precision, safeguard data privacy, and uphold fairness. Consistently validating and collaborating with domain experts helps reduce the chances of misidentifying dangers while following ethical norms for data collecting safeguards data privacy and ownership. Employing methods to identify and address biases guarantees fair and impartial outcomes among various species or groups. In addition, it is important for AI to enhance human expertise rather than substitute it, enabling taxonomists and ecologists to concentrate on more intricate assignments.

Moreover, it is imperative to avert possible abuse of AI technology, such as in illicit wildlife trafficking or habitat devastation, through the implementation of security protocols and cooperation with regulatory entities. It is important to take into account the ecological consequences of AI by optimizing algorithms to be more efficient and using renewable energy sources. Our objective is to ethically develop and implement AI for species identification by addressing ethical considerations. This involves maximizing the advantages while minimizing any potential harm.

VII. FUTURE WORK

We intend to include further features into our Model. At present, our focus lies on three distinct species: humans, plants, and animals. Shortly, we aim to develop a recognition system that encompasses all five kingdoms. Nevertheless, photographing micro species using a regular camera proves to be difficult; a specialized micro-enabled camera is required for image acquisition. There is potential for enhancing gender prediction accuracy and expanding the number of classifications within animal species. It is possible to forecast the familial organization of animal and plant categories. This Model can be seamlessly integrated with live scripting. In addition to automated species identification, we may also utilize speech recognition technology, which will provide us with an auditory output detailing the features of the species.

1) **Objective:** Enhance the precision of gender identification within specified species.
2) **Approach:** Enhance the dataset by including photos that are labeled with gender information and improve the model's structure to more effectively capture characteristics that are distinctive to gender.
3) **Challenges:** The degree of gender dimorphism differs greatly among species, necessitating customized strategies for distinct populations. Expanded Taxonomy within Animal Species.
4) **Objective:** Enhance the degree of classification by incorporating subspecies and family levels.
5) **Approch:** Employ hierarchical classification methodologies to augment the training dataset by incorporating images of subspecies and other families.
6) **Challenge:** Acquiring a heterogeneous and exhaustive dataset that encompasses different subspecies and families.Integration with live scripting and auditory outputs.
7) **Objective:** Effortlessly incorporate the model into live scripting to deliver instantaneous recognition and audible explanations.
8) **Approach:** Develop and incorporate real-time picture processing functionalities and combine them with text-to-speech systems to deliver aural responses.

9) **Challenges:** Ensuring immediate and precise execution and correctness, especially in different environmental circumstances.Technology for recognizing and interpreting spoken language.
Our objective is to greatly improve the capabilities and usefulness of our automated species identification system by focusing on these particular goals, approaches, and potential problems.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

AUTHORS CONTRIBUTION

All the authors contributed equally to this research work.

ACKNOWLEDGMENT

The authors acknowledge the support provided by the Department of Computer Science, Maynooth University, Ireland. Innovation Value Institute, Maynooth University, Ireland.

REFERENCES

[1] M. S. Norouzzadeh, D. Morris, S. Beery, N. Joshi, N. Jojic, and J. Clune, "A deep active learning system for species identification and counting in camera trap images," *Methods Ecol. Evol.*, vol. 12, no. 1, pp. 150–161, Jan. 2021.
[2] K. Frafjord, "Activity patterns of the soprano pipistrelle pipistrellus pygmaeus throughout the year in southern Norway," *BMC Zoology*, vol. 6, no. 1, pp. 1–8, Dec. 2021.
[3] P. H. W. Biedermann and F. E. Vega, "Ecology and evolution of insect–fungus mutualisms," *Annu. Rev. Entomol.*, vol. 65, no. 1, pp. 431–455, Jan. 2020.
[4] C.-H. Wu, W.-L. Wei, J.-C. Lin, and W.-Y. Lee, "Speaking effect removal on emotion recognition from facial expressions based on eigenface conversion," *IEEE Trans. Multimedia*, vol. 15, no. 8, pp. 1732–1744, Dec. 2013.
[5] M. Favorskaya and A. Pakhirka, "Animal species recognition in the wildlife based on muzzle and shape features using joint CNN," *Procedia Comput. Sci.*, vol. 159, pp. 933–942, Jan. 2019.
[6] R. Breuer and R. Kimmel, "A deep learning perspective on the origin of facial expressions," 2017, *arXiv:1705.01842*.
[7] Z. Zhang, L. Yin, Y. Peng, and D. Li, "A quick survey on large scale distributed deep learning systems," in *Proc. IEEE 24th Int. Conf. Parallel Distrib. Syst. (ICPADS)*, Dec. 2018, pp. 1052–1056.
[8] H. Jung, S. Lee, J. Yim, S. Park, and J. Kim, "Joint fine-tuning in deep neural networks for facial expression recognition," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 2983–2991.
[9] M. A. Iqbal, Z. Wang, Z. A. Ali, and S. Riaz, "Automatic fish species classification using deep convolutional neural networks," *Wireless Pers. Commun.*, vol. 116, no. 2, pp. 1043–1053, 2021.
[10] A. A. dos Santos and W. N. Gonçalves, "Improving Pantanal fish species recognition through taxonomic ranks in convolutional neural networks," *Ecological Informat.*, vol. 53, Sep. 2019, Art. no. 100977.
[11] M. Toğaçar, B. Ergen, and Z. Cömert, "Classification of flower species by using features extracted from the intersection of feature selection methods in convolutional neural network models," *Measurement*, vol. 158, Jul. 2020, Art. no. 107703.
[12] G. Saini, A. Khamparia, and A. K. Luhach, "Classification of plants using a convolutional neural network," in *Proc. 1st Int. Conf. Sustain. Technol. Comput. Intell.* Cham, Switzerland: Springer, 2020, pp. 551–561.
[13] J. Stastny, M. Munk, and L. Juranek, "Automatic bird species recognition based on birds vocalization," *EURASIP J. Audio, Speech, Music Process.*, vol. 2018, no. 1, pp. 1–7, Dec. 2018.
[14] V. Allken, N. O. Handegard, S. Rosen, T. Schreyeck, T. Mahiout, and K. Malde, "Fish species identification using a convolutional neural network trained on synthetic data," *ICES J. Mar. Sci.*, vol. 76, no. 1, pp. 342–349, Jan. 2019.
[15] S. Habib, N. U. Haq, and W. Younas, "Correlation analysis of the global activity limitation indicator with disability measures in adults under 65 suffering from spinal and hepatic health issues," *Pakistan J. Sci. Res.*, vol. 4, no. 1, pp. 48–60, 2024.
[16] I. Banerjee, Y. Ling, M. C. Chen, S. A. Hasan, C. P. Langlotz, N. Moradzadeh, B. Chapman, T. Amrhein, D. Mong, D. L. Rubin, O. Farri, and M. P. Lungren, "Comparative effectiveness of convolutional neural network (CNN) and recurrent neural network (RNN) architectures for radiology text report classification," *Artif. Intell. Med.*, vol. 97, pp. 79–88, 2019.
[17] S. A. Pearline, V. S. Kumar, and S. Harini, "A study on plant recognition using conventional image processing and deep learning approaches," *J. Intell. Fuzzy Syst.*, vol. 36, no. 3, pp. 1997–2004, Mar. 2019.
[18] G. Hu, X. Yang, Y. Zhang, and M. Wan, "Identification of tea leaf diseases by using an improved deep convolutional neural network," *Sustain. Comput., Informat. Syst.*, vol. 24, Dec. 2019, Art. no. 100353.
[19] X. Zhu, M. Zhu, and H. Ren, "Method of plant leaf recognition based on improved deep convolutional neural network," *Cogn. Syst. Res.*, vol. 52, pp. 223–233, Dec. 2018.
[20] J. Chen, Q. Liu, and L. Gao, "Visual tea leaf disease recognition using a convolutional neural network model," *Symmetry*, vol. 11, no. 3, p. 343, Mar. 2019.
[21] S. Zhang, S. Zhang, C. Zhang, X. Wang, and Y. Shi, "Cucumber leaf disease identification with global pooling dilated convolutional neural network," *Comput. Electron. Agricult.*, vol. 162, pp. 422–430, Jul. 2019.
[22] J. Shao and Y. Qian, "Three convolutional neural network models for facial expression recognition in the wild," *Neurocomputing*, vol. 355, pp. 82–92, Aug. 2019.
[23] S. Hossain, S. Umer, V. Asari, and R. K. Rout, "The deep learning-based facial expression recognition system is a unified framework for diversified applications," *Appl. Sci.*, vol. 11, p. 9174, Jan. 2021.
[24] H. Zhang, A. Jolfaei, and M. Alazab, "A face emotion recognition method using convolutional neural network and image edge computing," *IEEE Access*, vol. 7, pp. 159081–159089, 2019.
[25] D. Bisen, "Deep convolutional neural network based plant species recognition through features of leaf," *Multimedia Tools Appl.*, vol. 80, no. 4, pp. 6443–6456, Feb. 2021.
[26] D. Rathi, S. Jain, and S. Indu, "Underwater fish species classification using convolutional neural network and deep learning," in *Proc. 9th Int. Conf. Adv. Pattern Recognit. (ICAPR)*, Dec. 2017, pp. 1–6.
[27] X. Wang, C. Zhang, and S. Zhang, "Multiscale convolutional neural networks with attention for plant species recognition," *Comput. Intell. Neurosci.*, vol. 2021, no. 1, Jan. 2021, Art. no. 5529905.
[28] L. A. B. Prieto and Z. K. Oplatkova, "Emotion recognition using AutoEncoders and convolutional neural networks," *MENDEL*, vol. 24, no. 1, pp. 113–120, Jun. 2018.
[29] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16 × 16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.
[30] A. Manimaran, D. Chandramohan, S. G. Shrinivas, and N. Arulkumar, "A comprehensive novel model for network speech anomaly detection system using deep learning approach," *Int. J. Speech Technol.*, vol. 23, no. 2, pp. 305–313, Jun. 2020.
[31] A. Chakraborty, D. Kumer, and K. Deeba, "Plant leaf disease recognition using fastai image classification," presented at the *Proc. 5th Int. Conf. Comput. Methodologies Commun. (ICCMC)*, Apr. 2021.
[32] M. Brahimi, K. Boukhalfa, and A. Moussaoui, "Deep learning for tomato diseases: Classification and symptoms visualization," *Appl. Artif. Intell.*, vol. 31, no. 4, pp. 299–315, Apr. 2017.
[33] W. Zeng, H. Li, G. Hu, and D. Liang, "Lightweight dense-scale network (LDSNet) for corn leaf disease identification," *Comput. Electron. Agricult.*, vol. 197, Jun. 2022, Art. no. 106943.
[34] Y Library. (2021). *Yale University Library*. [Online]. Available: https://web.library.yale.edu/quicksearch/tips/databases
[35] Kaggle. (2019). *Kaggle Library*. [Online]. Available: https://www.kaggle.com/alessiocorrado99/animals10
[36] TNG Association. (2013). *Quantitative for Plants Datasets: Number of Images Concerning Groups, Family, and Species Name*. [Online]. Available: https://www.quantitative-plant.org/dataset/plant-database

SHAHEER HABIB received the Bachelor of Science degree in computer science from the University of Sargodha, Campus Mianwali, Mianwali, Pakistan, in 2019, and the Master of Science degree in computer science from the University of Lahore, Lahore Campus, in 2022. He is currently pursuing the Ph.D. degree with the University of Engineering and Technology. He is a Lecturer and the Head of the Department of Computer Science, Obaid Noor Institute of Medical Science, Mianwali. He is a Visiting Lecturer with the Department of Computer Science, University of Mianwali, Mianwali. He is a Ph.D. Scholar. He is interested in study areas, such as machine learning, computer vision, and medical imaging. He was awarded the bronze medal for the Bachelor of Science degree and the gold medal for the Master of Science degree.

MUBASHIR AHMAD received the B.Sc. degree (Hons.) from Hazara University Mansehra, Pakistan, the M.S.C.S. degree from IQRA University Islamabad, Pakistan, in 2012, and the Ph.D. degree from the Beijing Institute of Technology China, in 2019. From 2012 to 2014, he was a PHP Programmer with Microcom IT. He is currently an Assistant Professor with the Computer Science Department of Computer Science, COMSATS University Islamabad, Abbottabad Campus. His research interests include deep learning, pattern recognition, and medical imaging. He was a recipient of Chinese Government Scholarship for Ph.D. studies and got an Excellent Research Paper Award in IGTA-Springer, Beijing, China, in 2017. He also serves as a Reviewer for IEEE ACCESS.

YASIN UL HAQ received the B.S. degree in software engineering from the University of Sargodha, Pakistan, in 2013, and the M.Sc. and Ph.D. degrees in computer science from the University of Engineering and Technology, Lahore, Narowal, Pakistan, in 2017. He is currently a Lecturer with the Department of Computer Science and Engineering, University of Engineering and Technology, Lahore. His research interests include machine learning, remote sensing, web semantics, software engineering, big data, and cybersecurity.

RABIA SANA received the B.S. degree in software engineering from the University of Sargodha, in 2013, and the M.Sc. degree in computer science from Virtual University, in 2020. She is currently a Lecturer of computer science with the University of Engineering and Technology, Lahore, Narowal. Her research interests include software engineering, semantic web, SQA, information systems, machine learning, and database systems.

ASIA MUNEER received the Bachelor of Science degree in computer science from the University of Sargodha, Campus Mianwali, Mianwali, in 2017, and the Master of Science degree in computer science from the University of Lahore, Lahore, in 2020. She is currently a Lecturer with the Department of Computer Science and IT, Punjab Higher Education Department, Pakistan. Her research interests include information security, information systems, machine learning, and artificial intelligence.

MUHAMMAD WASEEM received the Bachelor of Science degree in computer science from the University of Sargodha, in 2009, and the Master of Science degree in computer science from the Virtual University of Pakistan, in 2021. He is currently a Lecturer with the Department of Computer Science, Punjab Higher Education Department, Pakistan. His research interests include machine learning, computer vision, and medical imaging.

MUHAMMAD SALMAN PATHAN received the Ph.D. degree from Beijing University of Technology (BJUT), China. He is currently an Assistant Professor with the School of Computing, Dublin City University. He has published several research articles in these areas. His research interests include machine learning, information security, and sustainability.

SOUMYABRATA DEV (Member, IEEE) received the B.Tech. degree (summa cum laude) from the National Institute of Technology Silchar, India, in 2010, and the Ph.D. degree from Nanyang Technological University (NTU), Singapore, in 2017. In 2015, he was a Visiting Student with the Audiovisual Communication Laboratory (LCAV), Ecole Polytechnique Federale de Lausanne (EPFL), Lausanne, Switzerland. He is currently an Assistant Professor with the School of Computer Science, University College Dublin, Ireland, and a SFI Funded Investigator with ADAPT SFI Research Centre, Dublin. He has authored more than 120 publications in leading journals and conferences. His research interests include remote sensing, statistical image processing, machine learning, and deep learning.