

# Identification of Indian Bird Species to Promote Conservation Endeavors

Mr. Yogesh K. Sharma<sup>1</sup>, Mandar S. Karyakarte<sup>1</sup>, Mr. Gajanan H. Chavhan<sup>1</sup>  
[yogesh.sharma@viit.ac.in](mailto:yogesh.sharma@viit.ac.in), [mandar.karyakarte@viit.ac.in](mailto:mandar.karyakarte@viit.ac.in), [gajanan.chavhan@viit.ac.in](mailto:gajanan.chavhan@viit.ac.in)

Dr. Rais Khan<sup>2</sup>, Mr. Milind Patil<sup>1</sup>, Mr. Rajendra S. Talware<sup>1</sup>  
[rais.khan@ghru.edu.in](mailto:rais.khan@ghru.edu.in), [milind.patil@viit.ac.in](mailto:milind.patil@viit.ac.in), [rajendra.talware@viit.ac.in](mailto:rajendra.talware@viit.ac.in)

Vishwakarma Institute of Information Technology, Pune, India<sup>1</sup>  
Engineering G H Rasoni University, Amravati, India<sup>2</sup>

## Article Info

Page Number: 8803 - 8824

Publication Issue:

Vol 71 No. 4 (2022)

## Abstract

Birds are vertebrate animals that are found all across the globe and are an integral part of most ecosystems. With a large variety of birds found globally and the significant role they play in conservation planning and decision-making, there is an increased interest shown by researchers, students as well as hobbyists to study them. Due to the many similarities among subspecies of birds, it becomes difficult to identify them and thus, a system that can help in the identification process is much desired. Using image classification techniques, bird identification systems can be developed which can be used to identify bird species by using their images. The current development in this domain is quite promising with many systems being developed for identifying birds using either their images, videos, or sounds, but most of them have been developed for North American birds. The purpose of our project is to develop a bird identification system specifically for Indian birds to provide a convenient identification tool to the students, researchers, and hobbyists in India. We make use of the GoogLeNet pre-trained classification model and apply transfer learning techniques to train the model on our dataset of Indian birds. The model yielded an accuracy of 99.51% on the testing set, one of the highest among the existing identification systems for Indian birds.

## Article History

Article Received: 08 July 2022

Revised: 20 August 2022

Accepted: 12 September 2022

Publication: 23 October 2022

**Keywords**— Birds, Image Classification, Deep Learning, Transfer Learning, GoogLeNet

## I. INTRODUCTION

Birds are one of the most important species due to their prominent role in most ecosystems and food chains [1]. A recent study [2] has estimated that there are more than 10,000 bird species and more than 50 billion birds across the world. Bird populations are good indicators of ecosystem health and trends in their population can be seen as an indicator of wider changes within their habitats [3]. It is due to this diversity and large population as well as the significant role they play in maintaining environments for plant distribution, agriculture and biological conservation that have made birds one of the most important parameters considered during conservation planning and environmental assessments. Hence, birds have been extensively studied by researchers (ornithological research) as well as by amateur observers [1]. These amateur observers are typically bird watchers who engage in bird watching as a recreational activity [4] often capturing images of the birds in the process. Identification of birds thus, is important as it can aid researchers, scholars as well as amateur observers in furthering their research and studies about birds and their surrounding environment.

Identification of bird species is a perplexing problem that is faced by both researchers as well as bird observers [4][5]. Birds can be segregated based on their colour, shape, beak size, etc. and sometimes different birds may share a few similar characteristics between them making it difficult to identify the species with certainty. Also, certain constraints from the observer's end such as location, distance and equipment used to identify birds forces the observer to recognize birds with the naked eye based on their physical characteristics which make this a tedious process [4]. Ornithologists as well as seasoned bird watchers have also found it difficult to identify bird species solely based on observed physical characteristics. Thus, there is a need for a system which can help in identification of bird species, reducing the time spent and effort required in doing the same identification manually.

Birds can be identified using either their captured images, recorded audio or video data. Audio and video data can be processed, and the signals can be analyzed for the identification process. However, due to the complexity introduced by the noise in audio data and the presence of other subjects in video data make it difficult to process them for further analysis. Since bird images can be easily captured and processed, they are the ideal candidates for identification of bird species [5]. Image classification techniques based on deep learning are used in such bird identification systems. Using a sufficiently large amount of image data, deep learning-based image classification algorithms can be trained and deployed for further use by researchers and bird watchers. Most

existing bird systems are deployed on a web based or mobile based platform making them widely accessible and easier to use. In this survey, the existing bird identification systems and their approaches and performance have been reviewed.

The following is how the remainder of this paper is structured: Section 2 briefly introduces image classification, its techniques and approach. In Section 3, the existing systems for bird identification have been reviewed. A summary and analysis of the survey is provided in section 4. We introduce our bird identification system in Section 5 with an analysis of its results in section 6. Finally, the paper is summarized, and the future scope is discussed in section 7.

## II. IMAGE CLASSIFICATION

Image classification is the process of classifying image pixels into a finite set of classes based on their data value. The image pixels are assigned to a particular class if they satisfy the rules to fit in that class. The classes of classification are known if the user can separate the classes based on the training data else the classes are unknown [6]. Image classification is a relatively complex process and can be affected by many factors [7]. Images can be hard to classify especially when they contain noise and background clutter or even when the image is of poor quality. Also, having more than one object in the image can make the classification process difficult [8]. Since image classification can be used to extract information

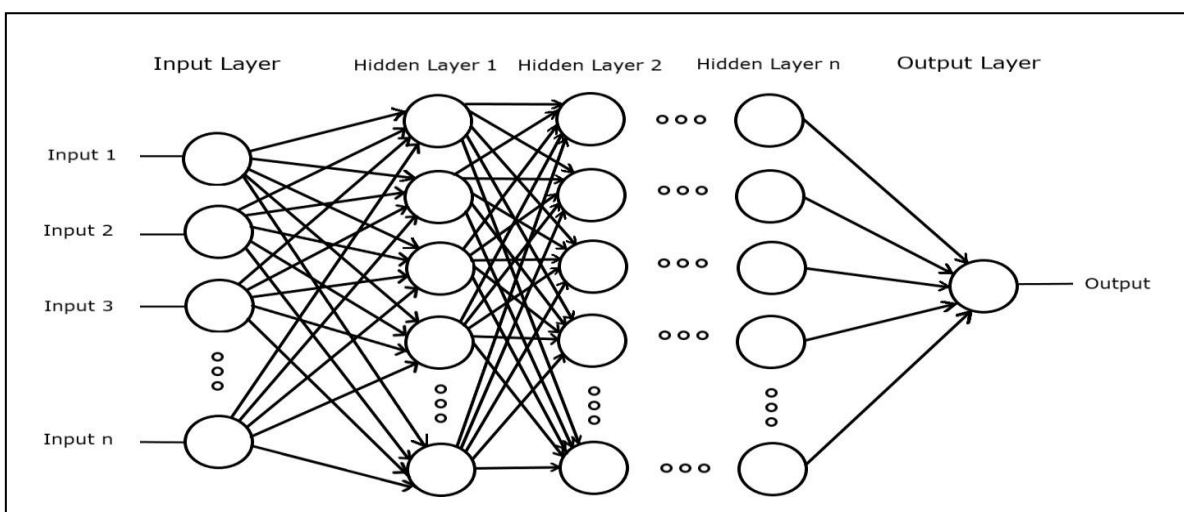


Fig. 1. Basic Structure of an ANN model

from digital images [6] and the characterization results can shape the reason for some ecological and financial applications, scientists have been chipping away at creating progressed arrangement approaches and strategies for further developing the grouping exactness [7]. Picture arrangement

methods can be grouped into administered and solo, or parametric and non-parametric, or hard and delicate classifiers.

## **A. *Image Classification Techniques***

There are many image classification techniques available today with new technologies continuously being developed by researchers. This section introduces some of these advanced image classification techniques that have been used in the surveyed bird classification systems.

### **1. *Artificial Neural Network***

Artificial Neural Networks (ANN) are non-parametric classifiers which basically have no assumption of the data and don't make use of statistical parameters to calculate class separation [6][7][8]. The design of the architecture of ANN is inspired from the design of the human nervous system. ANN at the basic level have two layers – input and output, but some systems may have additional layers called hidden layers as well which are connected to each other via weighted links [8]. ANN have two phases namely the training phase and the testing phase [6]. In the training phase, the ANN learns the patterns from the training data and stores it as knowledge. In the testing phase, the ANN uses the knowledge gained from training to predict the class of newly fed data. The main advantages of ANN are its high computation rate, data driven training and efficiency in dealing with noisy inputs [6]. The large time taken for training [6][8], poor semantics and problem of overfitting [6] are some of the disadvantages of ANN. The general performance of ANN is dependent on its network structure [7]. ANN can be used for image classification by texture feature extraction and then training the network using back propagation algorithm [9]. Fig. 1 shows the basic structure of an ANN model.

### **2. *Support Vector Machines***

Support Vector Machines (SVM) are binary, non- parametric classifiers [6][7][8] that separate the classes using a linear boundary [6][8][9]. SVM is a machine learning algorithm that is based on statistical learning theory and its goal is to select an optimal separating hyperplane - a plane in a multi- dimensional space that separates the data sample of two classes, that maximizes the margins from the closes data points to the plane [7][8][9]. Fig. 2 shows an example of an optimal hyperplane separating the data samples of two classes. SVM can also be used for multi-class classification but requires the use of some integration strategies [8][9]. The correct choice of kernel function in SVM is very important as it performs the mapping from the input space to the feature space [6][9]. The main advantages of SVM are its ability to generalize easily, not suffering from overfitting and use of non-linear transformation [6]. The time taken for training, less transparency

of results [8], complex algorithm design and difficulty in definition of optimal parameters [6] are some of the disadvantages of SVM. The general performance of an SVM is dependent on its kernel parameters [7][9].

### **B. Image Classification Approach**

An Image Classification task usually consists of many steps from capturing the images to training a model to assessing the model and deploying it. Fig. 3 shows the steps involved in the image classification approach.

#### **a) Capturing digital images**

Images that are to be used as part of the training and testing phases of an image classification algorithm are captured using cameras.

#### **b) Defining the classes**

Depending on the properties of the images and the classification objective, the classes of the images are defined. For supervised algorithms, the images must be segregated and labeled accordingly.

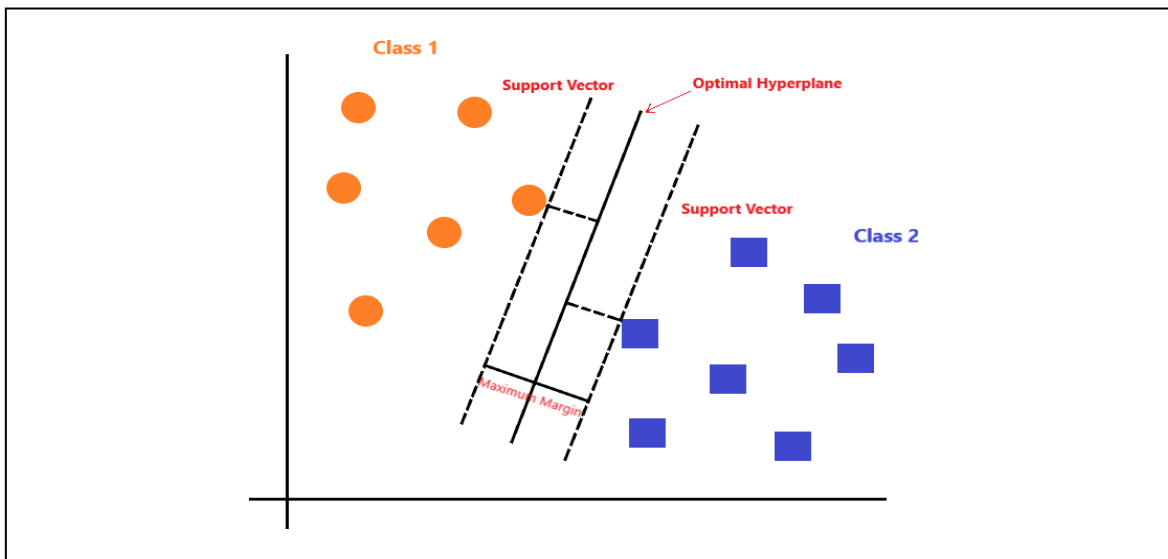


Fig. 2. Optimal hyperplane separating data samples of two classes

#### **c) Splitting of image data**

For most image classification algorithms, the image data is split into training, testing and validation sets. The splitting of the data is done to prevent overfitting and to accurately evaluate the model.

#### **d) Pre-processing of the image data**

The image data obtained via cameras may contain some background noise, additional objects, and may also not have uniform dimensions. To improve the training accuracy of the algorithm, pre-processing of the image data takes place where the images are resized, filters are applied to reduce

noise, channels are reduced (usually grayscale images are used) and normalization is applied.

*e) Feature extraction*

In feature extraction, the image pixels are mapped into the feature space. The meaningful features that characterize the image samples are extracted and stored in a feature vector. Using these extracted features, the similarity and difference between images can be determined which can be used to differentiate the defined image classes.

*f) Classification of images*

To create an image classification model that can be used for classification, an appropriate classification technique is first selected based on the available data and the classification objective. The model is then trained using the training data where the model learns to categorize the images into classes by using some strategies and optimization techniques.

*g) Evaluation of output and accuracy*

The result obtained from the classifier is verified and the accuracy of the model is calculated. Based on calculated accuracy and model performance, the model may be further improved or deployed to be used for further classification.

### III. LITERATURE SURVEY

Since images and videos are the most popular and reliable media for creating bird datasets and identifying them, the scope of this review has been restricted to systems that have only used bird images or videos for classification.

A. Varghese, et al. [4] proposed a deep learning platform for recognition of bird species using its images. The proposed method uses convolutional neural networks (CNN) for classification as well as prediction. The system also uses a skip connection oriented neural network for improving the feature extraction process. The Caltech-UCSD Birds dataset is used which is a fine-grained biological image classification dataset containing 11,788 images of 200 different categories of bird species. The dataset was split with more than 60% of the images allocated to the training set and the rest being randomly allocated to the test set and validation set for fine-tuning. The CNN models use stacks of convolutional layers which comprise an input layer, two fully connected (FC) layers and a final output softmax layer. Each set of convolutional layers comprised of 5\*5 convolution, batch normalization (BN), Rectified linear unit (ReLU) activation and pooling layers. The model also uses skip layer connections that offer an alternative path for the gradient, where the output of a layer is fed as the input to a layer later in the network skipping a few layers in between. This

improves feature extraction by using weighted summation of corresponding layers. An image of the bird to be identified is given as input to the system and the image is converted into grayscale and then into a matrix format. Different alignments of the image are then sent to the CNN for feature extraction. The feature vectors that are extracted from the image are then sent to the CNN with the trained model. The extracted features are then passed to the predictive model which compares the features with the test data and the species of the bird is then identified. The proposed system was evaluated using 100 bird images. It achieved an 100% accuracy while predicting images containing birds as bird images.

S. Raj, et al. [5] developed a web-based bird identification platform that uses deep learning. The system uses the CNN

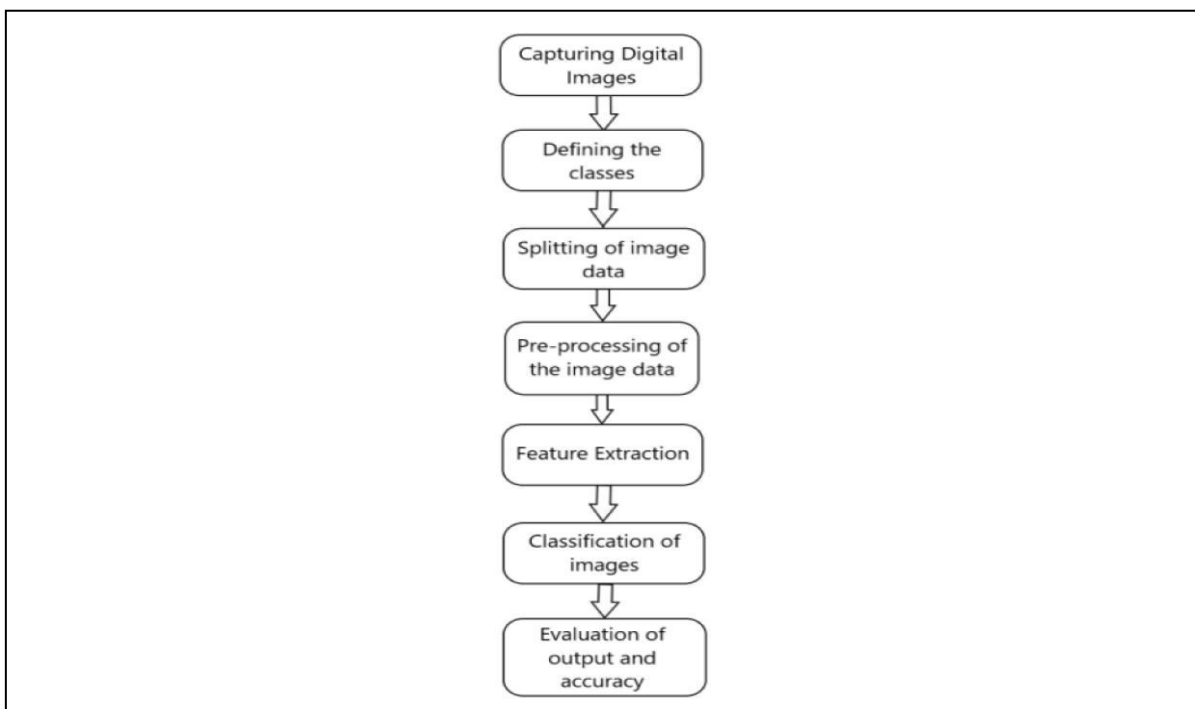


Fig. 3. Steps involved in image classification approach

deep learning algorithm for identification. The dataset was fabricated utilizing Microsoft's Bing Picture Search Programming interface and contained 8218 pictures of 60 types of birds saw as in the Asian sub-mainland. An irregular distribution of the pictures was finished with 80% of the pictures designated to the preparation set and the excess 20% dispensed to the testing set. The pictures used to prepare the organization had aspects of 96X96X3. The Picture Generator class was utilized for dataset expansion to expand the assortment of accessible data and forestall overfitting.

The CNN engineering utilized was a more modest, compact form of the VGGNet network which comprises of 3x3 convolutional layers stacked on one another at expanding profundity, max pooling layers that limit the size of the picture and the quantity of boundaries, and completely associated layers toward the finish of the organization before a softmax classifier. The carried out CNN engineering comprises of different Convolution and ReLU layers stacked together where the principal convolution block has a Convolution layer with 32 channels and a 3x3 component indicator. Then, the ReLU capability is applied which is trailed by a Pooling layer that consolidates a 3X3 pool to lessen the elements of the pictures from 96x96 to 32X32. The following convolution block in the stack has a Convolution layer with an expanded channel size of 64 with a similar 3x3 element locator. This is trailed by the ReLU capability and Max Pooling with a diminished pool window size of 2X2 at steps of 2. The last convolution block has a Convolution layer where channel size is expanded to 124, which is trailed by execution of ReLU capability. Max Pooling is applied again with a similar pool window size of 2X2 at steps of 2. To forestall overfitting, a Dropout Layer with a dropout worth of 0.25 is utilized. Then, the Completely Associated Layers are added utilizing a Thick Layer of size 1024. Another Dropout Layer with worth of 0.50 is carried out. At last, a Softmax classifier is utilized for anticipating a solitary class out of various classes. The conveyed CNN model was then prepared utilizing Adam Analyzer. The proposed system had a classification accuracy of 93.19% on the training set and an accuracy of 84.91% on the testing set.

J. Atanbori, et al. [10] proposed a computerized framework that is fit for ordering the types of individual birds during flight utilizing video information. The dataset utilized for the framework comprised of video groupings of flying birds from 7 distinct species which were recorded utilizing a Casio Exilim ZR100 at 240 edges each second. Each recorded species comprised of in excess of ten people, aside from Great Starlings, which had three. The birds' outline was removed utilizing the foundation Gaussian blend model. The shapes were gotten from the parallel picture utilizing a form calculation to distinguish the associated parts. Each outline was fitted with a situated bouncing box and measurements, for example, level, width, hypotenuse, centroid, outline and shape focuses were estimated. For any bird  $j$  followed all through  $n$  outlines, the direction model was characterized as the focal point of the fitted limited box (for example the centroid) and was given by the situation:

$$T_j = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\} \quad (1)$$

where  $T$  addresses the direction and  $x$  and  $y$  are the directions of the centroid. The direction, being uproarious, was smoothened by applying a crate channel with  $1 \times 3$  piece. The outlines were



divided and afterward the variety minutes, shape minutes, grayscale histogram, gabor channel and log-polar highlights were extricated. These highlights were linked to frame one component vector for order of bird species by variety, shape and surface. For the directions and fitted bends, the ebb and flow scale space (CSS), turn based highlights, centroid distance, area and arch in light of sine and cosine highlights were separated. These highlights along with wing beat frequencies were connected to shape one component vector for grouping of bird species by their directions. The features were represented as statistical features to provide information on the location, variability

type	patch size/ stride	output size	depth	#1×1	#3×3 reduce	#3×3	#5×5 reduce	#5×5	pool proj	params	ops
convolution	7×7/2	112×112×64	1							2.7K	34M
max pool	3×3/2	56×56×64	0								
convolution	3×3/1	56×56×192	2		64	192				112K	360M
max pool	3×3/2	28×28×192	0								
inception (3a)		28×28×256	2	64	96	128	16	32	32	159K	128M
inception (3b)		28×28×480	2	128	128	192	32	96	64	380K	304M
max pool	3×3/2	14×14×480	0								
inception (4a)		14×14×512	2	192	96	208	16	48	64	364K	73M
inception (4b)		14×14×512	2	160	112	224	24	64	64	437K	88M
inception (4c)		14×14×512	2	128	128	256	24	64	64	463K	100M
inception (4d)		14×14×528	2	112	144	288	32	64	64	580K	119M
inception (4e)		14×14×832	2	256	160	320	32	128	128	840K	170M
max pool	3×3/2	7×7×832	0								
inception (5a)		7×7×832	2	256	160	320	32	128	128	1072K	54M
inception (5b)		7×7×1024	2	384	192	384	48	128	128	1388K	71M
avg pool	7×7/1	1×1×1024	0								
dropout (40%)		1×1×1024	0								
linear		1×1×1000	1							1000K	1M
softmax		1×1×1000	0								

Fig. 4. GoogLeNet's Inception Incarnation

and appearance of the distribution of data and ensure that classification can be performed in real time. The measurable highlights figured incorporate the mean, standard deviation, skewness, kurtosis, energy, entropy, greatest, least, neighborhood maxima, nearby minima and number of zero intersections. The appearance and movement include sets were tried utilizing an Ordinary Bayes classifier and a Help Vector Machine and assessed freely. For assessment, a straightforward cross-approval conspire in view of a 70% preparation set, and 30% test set was utilized. The assessment was rehashed for four different test sets and found the middle value of to acquire the outcomes. For each trial run, the singular picture outlines from the preparation and test set were examined. A normal of 16,400 picture outlines from each preparing set, and a normal of 7,221 from each test set was utilized. The whole dataset included 162 recordings, each containing somewhere in the range of 0.25 and 5 seconds of fast video. The Ordinary Bayes classifier expected a Gaussian

combination model over the entire preparation information circulation, one part for every class, and assessed boundaries from the preparation information. The SVM classifier was carried out utilizing a spiral premise capability portion, with the gamma and cost boundaries enhanced utilizing a 5-crease framework look for definition and approval. Since a total matrix search is tedious, a coarse lattice search was performed first and after a decent locale on the network was found, a better framework search with that district was performed. The preparation time for appearance highlights utilizing Typical Bayes Classifier was 157.51 seconds and for a SVM classifier was 152.10 seconds. The preparation time for movement highlights utilizing Typical Bayes Classifier was 202.65 seconds and for a SVM classifier was 202.72 seconds. From the four tests performed with the Typical Bayes Classifier on the appearance highlight sets, the greatest characterization rate was 95% and the base was 89% with a normal of 92%. The best order was gotten for House Martins with 97% and the most minimal was for Cockatiels with 84%. For the SVM Classifier, the overall classification rate was 89% with the best classification obtained by Green Budgies with 94% and the lowest obtained by Cockatiels with 85%. For classification using the motion feature set, the overall classification rate was 37% based on four tests with the highest being 41% and the lowest being 33% among the tests. The highest classification rate was obtained for Green Budgies with 66% and the lowest rate was obtained for Black Birds with 8%. Using Normal Bayes Classifier, the overall classification rate for the majority of the species was above 40% indicating that the motion features have the ability to provide additional differentiation. In the proposed system, the appearance and motion features were used separately for classification. However, these features could be combined to create a more robust species classification system.

**B.Qiao, et al. [11]** proposed another bird species calculation in light of a multi — scale choice tree and the SVM system to recognize bird species. For the proposed framework, the variety second, the exceptional variety, the shade of the head, the component of snout and second invariants are incorporated as an element to portray the bird. Most accessible pictures of birds are generally their side profiles thus the noses of the birds are effectively noticeable. Since the nose highlights of birds are different across various species, it very well may be utilized to effectively recognize birds. As the size of pictures are not consistent, the bill qualities should take on a general worth thus the proportion of the level of the snout and the width of the mouth (R-HBWB) were viewed as in before strategies. Nonetheless, since levels of the noses change a considerable amount, the proposed framework utilizes the proportion of the distance of the eye to the base of the mouth and the width of the bill (R-ERWB) as it was a more successful component. Hu second invariants were utilized to portray the shape highlight and the angle proportion and thickness of the pictures were

likewise separated and added to the shape include. The unique variety element of the birds was found by picking red and blue as reference and working out every part of the bird sectioned from the foundation in the HSV space. The bird is red or blue if more than 8% of the proportion of the bird segmented from the background is occupied by red or blue pixels and if the bird has no red or blue, it will be divided into no special color. Since color statistical features are insensitive to changes in posture of the birds and can reflect the whole color characteristic of the image, it was used as part of the feature set.

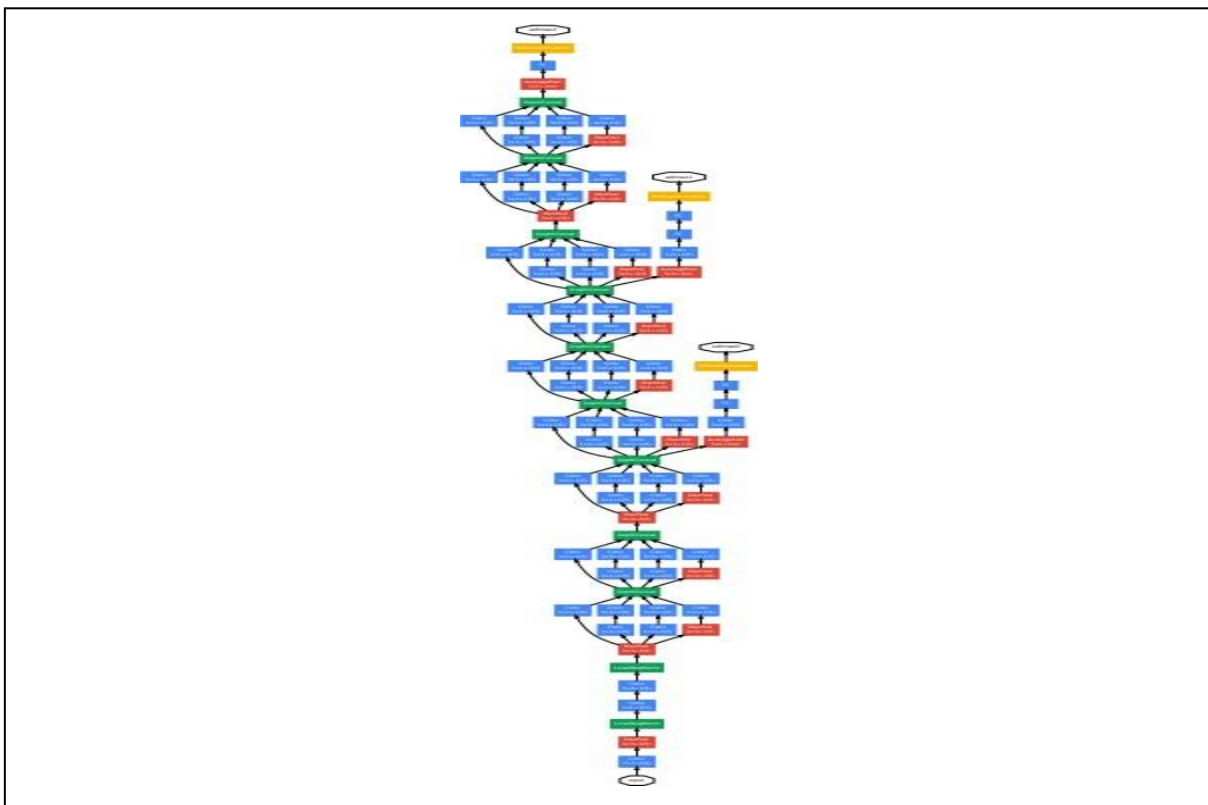


Fig.

### 5. GoogLeNet Network

The variety normal for the bird is addressed by the determined variety minutes. The main second, second and the third snapshot of three parts of the HSV variety space are extricated as variety measurable highlights on the grounds that the majority of the variety data is simply connected with the lower request second. The primary second mirrors the mean worth of three parts of the picture and the subsequent second and the third second mirror the difference in variety. As various types of bird heads might have various tones, the normal head tone is picked as the head include. The pixels around the bird's eyes are looked to get the head tone. On the off chance that the extent of any tone was over 10%, the variety was set apart as a component to perceive the bird. The proposed

framework was assessed on the Caltech bird's dataset (Whelp 200-2011) which contains 11788 pictures from 200 bird species in North America with roughly 60 pictures for every species and each picture being clarified with a bounding box around the bird. For the framework, 15 bird species were looked over the Whelp 200-2011 dataset with 20% of the quantity of pictures utilized for testing and different pictures utilized for preparing. The characterization rate was 83.87% when every one of the highlights (variety second, unique tone, head tone, bill element and second invariants) are utilized with the choice tree. The utilization of choice trees emphatically affected the arrangement rate, giving an increment of 3.23% in the characterization precision rate. It was likewise seen that when the list of capabilities without R-ERWB is utilized, the exactness rate comes around no less than 8% and when the list of capabilities without R-

HBWB is utilized, the exactness rate drops by 2%. Among the 3 variety spaces of RGB, YUV and HSV, it was noticed that variety highlights in the HSV variety space can more readily address the qualities with a pace of 75.26%.

M. Tayal, et al. [12] built a software solution to identify bird species using the concept of transfer learning. The application used the pre-trained AlexNet model for feature extraction and an SVM classifier for image classification. For the dataset, the first 200 image results for each bird species were downloaded using a Google Chrome extension. The images were then segregated into different folders and the folder names were used as labels during training and testing. The AlexNet model accepts images having  $227*227$  pixels and generally requires 100-200 images of the same class to better learn the features. The dataset was processed accordingly to meet these requirements. The images were compressed in bulk using the Caesium software and the dataset was then split into training and testing sets. Using transfer learning, the pre-trained AlexNet model was fine-tuned to identify the bird species. The extracted features were then used as predictor variables to fit a multiclass SVM classifier. This was implemented using the `fitcecoc` function that is a statistics and machine learning toolbox. The system was first trained using the training set and the accuracy was calculated. After that, the user can provide input to the system and the system returns the predicted label of the input image as the output after processing the image. The accuracy of the system was calculated for 4 bird species with a total of 40 testing images and had an accuracy of 85%. The highest accuracy was of 100% (Coppersmith Barbet) and the lowest was of 90% (Common Kingfisher).

A. Singh, et al. [13] developed a deep learning platform to assist users in recognizing bird species

using image recognition. In the proposed system, the user could upload an image to the system and the image will be stored in the database if it is not available in the dataset. The image would then be fed to the trained CNN model and the features such as the face, expression, angle, beak, etc. would be extracted from the image. In the model, some image processing is performed, the first step of which is acquisition. Here, the image is scaled and color conversion (RGB to Grayscale and vice versa) is performed. In the next step, some image enhancement is done which is used to extract the hidden details from the image. This is followed by image restoration, color image processing, and wavelets and multi-resolution processing. The next step is image compression which mainly deals with the image size or resolution. Morphological processing then extracts the image components that are useful in representation and description of shape. The image is then segmented into its constituent parts or objects after which a representation and description is chosen for transforming raw data into processed data. The classifier would then classify and predict the bird species using the extracted features and the trained dataset. The accuracy of the model using the training set was 93% and using the testing set of nearly 1000 images was 80%.

Omkarini, et al. [14] presented an automated model based on deep neural networks to identify the species of a bird given as the test dataset. The dataset that was used was the fine-grained Caltech-UCSD Birds (200-2011) dataset. The dataset contains 11,788 images representing 200 different bird species. This dataset was divided into training and testing sets where the training set received more than 60% of the total data and the testing set received the remaining. The images were then scaled to the same size ratio and converted to image pixel arrays using the cv2 library. The image pixel values were then normalized to reduce the noise and disturbances in the images. A CNN model was used for the identification of the images and was trained on the processed training dataset. The model yielded an accuracy of 98% on the training set. The trained model was then deployed using a client-server architecture. The user could upload an image to the system which is used as an input to the trained model. The model then gives the predicted bird species as an output which is finally displayed to the user.

**B.** Adjust, et al. [15] executed three unique strategies for bird order and contrasted their exhibitions with track down the best technique among them. With the end goal of the review, the Caltech-UCSD Birds (200-2011) dataset was utilized which incorporates a sum of 11,788 pictures across 200 types of birds alongside other data, for example, named noticeable bird parts, paired credits, and bouncing boxes encompassing the birds. The primary execution was a softmax relapse on the

paired properties which doesn't utilize PC vision yet is utilized as a standard for bird order. The highlights utilized for this technique were 312 bird highlights signifying ascribes like bill shape, wing tone, tail example, and generally speaking bird size, physically grouped by analysts at Caltech and UCSD. Most qualities were paired and the non-parallel ascribes like wing tone were broken into explicit wing colors like blue wing, dark wing, blue wing, and so on with the goal that they can have a twofold portrayal. Every one of the characteristics were then addressed involving a  $312 \times 1$  element vector for each preparing test. A regularized softmax calculation is first applied on the preparation information with explicitly ordered species where every subspecies is in an alternate class giving 200 distinct classes. The calculation is again applied on the preparation information with extensively grouped species where all subspecies are presently in a similar class bringing about 71 distinct classes. The precision on the explicitly characterized information was 54% and on the comprehensively ordered information was 70% because of the common credits between firmly related bird species in a similar class. The following execution was a multi-class SVM on Hoard and RGB highlights from the pictures. The contributions to this technique are the  $160 \times 160$  RGB resized pictures taken from the Caltech-UCSD Birds (200-2011) dataset. The component vectors are the histogram of situated angles (Hoard) linked with the RGB histogram values. Swines are highlight descriptors that can recognize outlines and invariance to mathematical and photometric changes with the exception of item direction. RGB histograms are utilized in the arrangement since the shade of birds is vital for distinguishing proof. The component vectors are then taken care of into a multiclass SVM which yields the expectation of the types of the bird. Pictures are parted into the preparation and testing set in light of the suggestions by the Fledgling dataset creators and this outcomes in a precision of simply 5% because of the confusion of the highlights of the bird brought about by the foundation. Be that as it may, in the wake of trimming the first pictures to their bouncing boxes, resizing them to  $160 \times 160$  pixels, concealing the foundation utilizing divisions given by the creators and afterward removing the Hoard and RGB elements and taking care of into a multi-class SVM, the exactness at long last increments to 9%. In the third and last execution, move learning is applied where a pre-prepared CNN with its last three layers changed is utilized for characterization. During introductory prototyping with the Keras library, a worked on brain network with two convolutional layers and two completely associated layers was tried with contributions of  $32 \times 32$  RGB pictures from the dataset yielding an exactness of 6% without tuning. Then, two techniques were tried utilizing the pre-prepared AlexNet model having 5 convolutional layers and 3 completely associated layers and prepared on the ImageNet dataset having more than 1,000,000 pictures. The information sources were  $227 \times 227$  RGB pictures from the Caltech-UCSD

Birds (200-2011) dataset.

The principal strategy utilizes the AlexNet model as a proper component extractor where the highlights from the last completely associated layer are separated and these elements are then taken care of into a multi-class SVM classifier which is utilized to foresee the test set. This technique yields a precision of 46%. In the subsequent technique, calibrating is utilized where the initial 22 layers of the pre-prepared AlexNet model are imported and the last three layers are supplanted with a completely associated layer, a softmax layer and an order yield layer. The learning rate, which was at first set to  $1e-4$ , was tuned to a lot bigger degree in the new layers when contrasted with the past layers to accelerate the model preparation. This strategy likewise yields a precision of 46% yet is computationally more costly since every one of the loads in the organization will be changed utilizing backpropagation, which consumes a large chunk of the day. Additionally, since the ImageNet dataset as of now contains pictures of birds, a portion of the elements are like the Whelp birds dataset, thus to lessen overfitting and cut down on the computational expense, the decent component extractor strategy is by all accounts the better choice. We have considered, adding to a total of 2031 images

Species	% Accuracy	Species	% Accuracy
BLACK-HEADED IBIS	100.00	GREY-FRONTED PIGEON	100.00
COMMON HAWK- CUCKOO	100.00	HOUSE CROW	100.00
COMMON WOOD PIGEON	96.55	INDIAN PEA FOWL	100.00
COPPERSMITH BARBET	100.00	PAINTED STORK	100.00
GREAT INDIAN BUSTARD	100.00	ROCK DOVE	100.00
GREATER COUCAL	100.00	SPECKLED WOOD PIGEON	100.00

#### IV. PROPOSED METHODOLOGY

The development of the proposed identification system for Indian birds can be divided into the following sections.

##### A. Creating the dataset

Since our project focuses on the identification of Indian birds, a suitable dataset of Indian bird images was needed for training and testing the model. Currently, no such dataset of Indian bird images is openly available for use and hence we created our own dataset for the purpose of this project. The dataset was created by scraping images from Google with around 120-140 images for each of the 15 bird species

images were then cleaned to have only the bird's body in the image and was resized to 224\*224 pixels. The dataset was then split into an 80:20 training and testing set using the splitfolders package in Python.

##### A. *Setting up the pre-trained model*

We have used the GoogLeNet model proposed by Christian Szegedy, et al. [16] as part of the ILSVRC14 competition. The model uses a deeper and wider Inception network and is 22 layers deep when only the layers with parameters are considered. The GoogLeNet network was trained on the ImageNet dataset having around 1.2 million images for training, 50,000 images for validation and 100,000 images for testing where each image is associated with a single ground truth category. The GoogLeNet architecture is as seen in Fig.4 and Fig. 5.

We make use of the pre-trained GoogLeNet network for our problem statement by fine-tuning it using transfer learning. The model was downloaded using the PyTorch framework. Before fine-tuning the model, we made some changes to the network so that it better suits our problem. The last fully connected layer of the GoogLeNet network is replaced with a sequential model having two linear layers where the first layer has 2048 in\_features and 1024 out\_features, and the second layer has 1024 in\_features and 15 out\_features.

##### B. *Fine-tuning the network*

After configuring the network and modifying the last layer to better suit our needs, we now fine-tune the network on our dataset of Indian Birds. First, some transformations such as random horizontal flips, normalization, resizing to 229\*229 pixels and then converting the image to tensor was applied to each of the images. The model was then trained on the training set for 20 epochs and a batch size of 32 on a Tesla K80 GPU. We used CrossEntropyLoss as the loss function and



Stochastic Gradient Descent as the optimizer with an initial learning rate of 0.001 and a momentum of 0.9.

## V. RESULTS

After training the model, the model was then tested using the testing set. The transformations that were applied on the testing set were applied on the training set as well. Experimental results with the testing set showed that the model has achieved an accuracy of 99.51% - one of the highest recorded on a dataset for Indian Birds. The accuracy for each class is as shown in Table 1.

## VI. CONCLUSION AND FUTURE SCOPE

In this paper, we have proposed a system for identification of Indian birds that uses transfer learning to fine-tune the pre-trained GoogLeNet model. We began by giving a brief introduction about the importance of birds and the importance of such a system especially in promoting conservation and helping spread knowledge about the different birds around us. We then introduced some concepts about image classification and reviewed the different existing bird identification systems, the image classification methods they have used, and the techniques needed for building such a system. Through this study we realized that there is a scope to create systems that can identify bird species to a greater accuracy, especially for Indian birds, which motivated us to propose our system.

The presented approach can still be developed further by including more bird species in the dataset to increase the range of birds that can be identified. Also, the system can be modified in such a way that it should be able to identify moving birds in videos as well. The system currently identifies only one bird in the image, but it can be modified so that multiple birds can be detected, identified and labelled in the same image.

## REFERENCES

- [1] M. Tabur and Y. Avyaz, "Ecological Importance of Birds", in *Second International Symposium on Sustainable Development*, Sarajevo, Bosnia and Herzegovina, June 8-9 2010, pp. 560-565
- [2] C. T. Callaghan, S. Nakagawa and W. K. Cornwell, "Global abundance estimates for 9,700 bird species", *Proceedings of the National Academy of Sciences*, vol 118, no 21, bl e2023170118, 2021.
- [3] P. J. Harrison, S. T. Buckland, Y. Yuan, D. A. Elston, M. J. Brewer, A. Johnston and J. W. Pearce-Higgins, "Assessing trends in biodiversity over space and time using the example of

British breeding birds”, *Journal of Applied Ecology*, 2014, vol. 51, no. 6, pp. 1650-1660, doi:  
<https://doi.org/10.1111/1365-2664.12316>

- [4] A. Varghese, ShyamKrishna K. and Rajeswari M., "Identification of Bird Species using a Deep Learning Technology", International
- [5] *Journal of Creative Research Thoughts (IJCRT)*, ISSN:2320-2882, Volume.9, Issue 1, pp.3743-3748, January 2021, Available: <http://www.ijcrt.org/papers/IJCRT2101459.pdf>
- [6] S. Raj, S. Garyali , S. Kumar and S. Shidnal, “Image based Bird Species Identification using Convolutional Neural Network”, *International Journal of Engineering Research & Technology (IJERT)*, June 2020, vol. 09, issue 06,
- [7] N. Thakur and D. Maheshwari, “A Review of Image Classification Techniques”, *International Research Journal of Engineering and Technology (IRJET)*, November 2017, vol. 04, issue 11
- [8] P. Gavali and J. S. Banu, “Deep Convolutional Neural Network for Image Classification on CUDA Platform,” *Deep Learning and Parallel Computing Environment for Bioengineering Systems*, pp. 99–122, 2019.
- [9] M. Jain and P. S. Tomar, “Review of Image Classification Methods and Techniques”, *International Journal of Engineering Research & Technology (IJERT)*, August 2013, vol. 02, issue 08
- [10] M. Seetha, I. V. Muralikrishna, B. L. Deekshatulu and B. L. Malleswari, "Artificial Neural Networks and Other Methods of Image Classification", *Journal of Theoretical and Applied Information Technology (JATIT)*, pp. 1039-1053, 2013.
- [11] J. Atanbori, W. Duan, J. Murray, K. Appiah and P. Dickinson, “Automatic classification of flying bird species using computer vision techniques”, *Pattern Recognition Letters*, vol. 81, pages - 53–62, 2016.
- [12] B. Qiao, Z. Zhou, H. Yang and J. Cao, "Bird species recognition based on SVM classifier and decision tree," 2017 First International Conference on Electronics Instrumentation & Information Systems (EIIS), 2017, pp. 1-4, doi: 10.1109/EIIS.2017.8298548.
- [13] M. A. Tayal, A. Mangrulkar, P. Waldey and C. Dangra, “Bird Identification using Image Recognition”, *HELIX*, vol. 08, no. 6, 2018
- [14] Image based Bird Species Identification
- [15] A. Singh, A. Jain and B. K. Rai, "Image based Bird Species identification", *International Journal of Research in Engineering, IT and Social Sciences*, April 2020, vol. 10, issue 04

- [16] V. Omkarini and G. K. Mohan, "Automated Bird Species Identification Using Neural Networks", *Annals of R.S.C.B.*, May 2021, vol. 25, issue 06
- [17] A.L. Alter and K. Wang, "An Exploration of Computer Vision Techniques for Bird Species Classification", 2017
- [18] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich, "Going deeper with convolutions," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1-9, doi: 10.1109/CVPR.2015.7298594.
- [19] Mulani, Altaf O., and P. B. Mane. "Watermarking and cryptography based image authentication on reconfigurable platform." *Bulletin of Electrical Engineering and Informatics* 6.2 (2017): 181-187.
- [20] Jadhav, Makrand M. "Machine Learning based Autonomous Fire Combat Turret." *Turkish Journal of Computer and Mathematics Education (TURCOMAT)* 12.2 (2021): 2372-2381.
- [21] Swami, Shweta S., and Altaf O. Mulani. "An efficient FPGA implementation of discrete wavelet transform for image compression." 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS). IEEE, 2017.
- [22] Shinde, Ganesh, and Altaaf Mulani. "A robust digital image watermarking using DWT-PCA." *International Journal of Innovations in Engineering Research and Technology* 6.4 (2019): 1-7.
- [23] Kulkarni, Priyanka R., Altaaf O. Mulani, and P. B. Mane. "Robust invisible watermarking for image authentication." *Emerging Trends in Electrical, Communications and Information Technologies*. Springer, Singapore, 2017. 193-200.
- [24] Bhanudas Gadade and Altaf Mulani, "Automatic System for Car Health Monitoring", *International Journal of Innovations in Engineering Research and Technology*, 57–62, 2022
- [25] Pratima Amol Kalyankar, Altaf O. Mulani, Sampada P. Thigale, Pranali Gajanan Chavhan and Makarand M. Jadhav, "Scalable face image retrieval using AESC technique", *Journal Of Algebraic Statistics* Volume 13, No. 3, p. 173 – 176, 2022
- [26] A. O. Mulani and G. N. Shinde, "An approach for robust digital image watermarking using DWT- PCA", *Journal of Science and Technology*, Vol.6, Special Issue 1, 2021 DOI: <https://doi.org/10.46243/jst.2021.v6.i04.pp59-62>
- [27] U. P. Nagane and A. O. Mulani, "Moving Object Detection and Tracking Using Matlab", *Journal of Science and Technology*, Vol.6, Special Issue 1, 2021 DOI: <https://doi.org/10.46243/jst.2021.v6.i04.pp63-66>

- [28] Priyanka Kulkarni and A. O. Mulani, "Robust Invisible Digital Image Watermarking using Discrete Wavelet Transform", International Journal of Engineering Research & Technology (IJERT), Vol. 4 Issue01, pp.139- 141, Jan.2015
- [29] Mulani, Altaf O., and Pradeep B. Mane. "High-Speed Area-Efficient Implementation of AES Algorithm on Reconfigurable Platform." Computer and Network Security (2019): 119.
- [30] Deshpande, Hrushikesh S., Kailash J. Karande, and Altaaf O. Mulani. "Area optimized implementation of AES algorithm on FPGA." 2015 International Conference on Communications and Signal Processing (ICCSP). IEEE, 2015.
- [31] Godse, A. P., and A. O. Mulani. Embedded systems. Technical Publications, 2009.
- [32] Mulani, Altaf O., and P. Mane. "Secure and area efficient implementation of digital image watermarking on reconfigurable platform." Int. J. Innov. Technol. Explor. Eng.(IJITEE) 8.2 (2018): 1.
- [33] Rahul G. Ghodake and A. O. Mulani, "Microcontroller Based Drip Irrigation System", Techno-societal 2016, International conference on advanced technologies for societal applications, pp. 109–115.
- [34] Amruta Mandwale and A. O. Mulani, "Different Approaches For Implementation of Viterbi decoder", IEEE International Conference on Pervasive Computing (ICPC), Jan. 2015.
- [35] Amruta Mandwale and A. O. Mulani, "Implementation of Convolutional Encoder & Different Approaches for Viterbi Decoder", IEEE International Conference on Communications, Signal Processing Computing and Information technologies, Dec. 2014.
- [36] Amruta Mandwale and A. O. Mulani, "Implementation of High Speed Viterbi Decoder using FPGA", International Journal of Engineering Research & Technology (IJERT), Feb. 2016
- [37] D. M. Korake and A. O. Mulani, "Design of Computer/Laptop Independent Data transfer system from one USB flash drive to another using ARM11 processor", International Journal of Science, Engineering and Technology Research, 2016.
- [38] Rahul G. Ghodake and A. O. Mulani, "Sensor Based Automatic Drip Irrigation System", Journal for Research, 53-56, 2016.
- [39] Rahul Shinde and A. O. Mulani, "Analysis of Biomedical Image", International Journal on Recent & Innovative trend in technology (IJRITT), July 2015
- [40] Rahul Shinde and A. O. Mulani, "Analysis of Biomedical Image using Wavelet Transform", International Journal of Innovations in Engineering Research and Technology (IJIERT), July 2015
- [41] A. O. Mulani and P. B. Mane, "Area optimization of cryptographic algorithm on less dense

reconfigurable platform,"2014 International Conference on Smart Structures and Systems (ICSSS), Chennai, 2014, pp. 86- 89

- [42] A.O.Mulani, M. M. Jadhav and Mahesh Seth, "Painless Non- invasive blood glucose concentration level estimation using PCA and machine learning" in the CRC Book entitled Artificial Intelligence, Internet of Things (IoT) and Smart Materials for Energy Applications, 2022.
- [43] Kamble, Akshata, and A. O. Mulani. "Google Assistant based Device Control." *Int. J. of Aquatic Science* 13.1 (2022): 550-555.
- [44] Pathan, Atik N., et al. "Hand Gesture Controlled Robotic System." *Int. J. of Aquatic Science* 13.1 (2022): 487-493.
- [45] Kolekar, Supriya D., et al. "Password Based Door Lock System." *Int. J. of Aquatic Science* 13.1 (2022): 494-501.
- [46] Swapnil Takale, Dr. Altaaf Mulani, "Video Watermarking System", *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, Volume 10, Issue III, Mar-2022.
- [47] J. P. Patale et al. "Python Algorithm to Estimate Range of Electrical Vehicle", *Telematique*, Volume 21, No. 1, 2022.
- [48] Jayshri Prakash Patale, A. B. Jagadale, A. O. Mulani, and Anjali Pise. "A Systematic Survey on Estimation of Electrical Vehicle". *Journal of Electronics, Computer Networking and Applied Mathematics (JECNAM)* ISSN : 2799-1156, vol. 3, no. 01, Dec. 2022, pp. 1-6, doi:10.55529/jecnam.31.1.6.
- [49] Kashid, M.M., Karande, K.J., Mulani, A.O. (2022). IoT-Based Environmental Parameter Monitoring Using Machine Learning Approach. In: Kumar, A., Ghinea, G., Merugu, S., Hashimoto, T. (eds) *Proceedings of the International Conference on Cognitive and Intelligent Computing. Cognitive Science and Technology*. Springer, Singapore. [https://doi.org/10.1007/978-981-19-2350-0\\_5](https://doi.org/10.1007/978-981-19-2350-0_5)
- [50] Swapnil Takale, and Dr. Altaaf Mulani. "DWT-PCA Based Video Watermarking". *Journal of Electronics, Computer Networking and Applied Mathematics (JECNAM)* ISSN : 2799-1156, vol. 2, no. 06, Nov. 2022, pp. 1-7, doi:10.55529/jecnam.26.1.7.
- [51] A. O. Mulani and Dr. P. B. Mane, "High throughput area efficient FPGA implementation of AES Algorithm", in the *Intech Open Access Book* entitled *Computer and Network Security*, Feb. 2019.
- [52] Hemlata M. Jadhav, Altaf Mulani and Makarand M. Jadhav, "Design and Development of

Chatbot based on Reinforcement Learning”, in the Wiley-IEEE book entitled Natural Language Processing using Machine and Deep Learning, 2022

- [53] V. B. Utpat, Dr. K. J. Karande, Dr. A. O. Mulani, “Grading of Pomegranate Using Quality Analysis”, International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 10 Issue II Feb 2022.