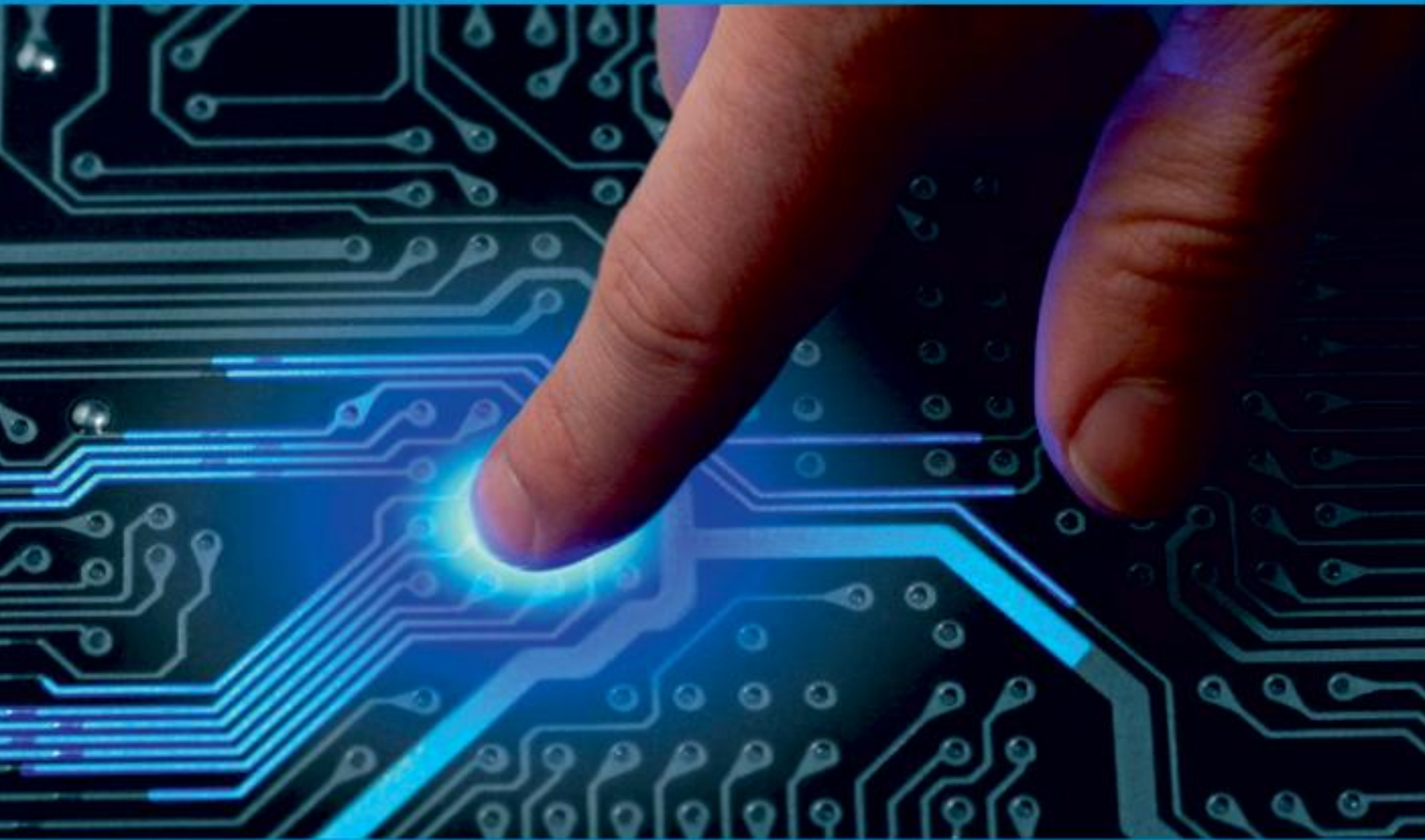
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 4, April 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104114 |

Bird Species Identification Using Audio

Shilpali Bansu, Nikhil Gosavi, Shreyas Bangar, Samruddhi Sawant, Shreya Gurav

Department of Computer Engineering, A C Patil College of Engineering, Kharghar, Navi Mumbai, India.

ABSTRACT: In this work, methods for identifying different bird species were researched, and automatic bird species recognition systems were established. Automatically recognizing bird cries without physical interaction has been a significant and tedious task for major studies in several subfields of taxonomy and other ornithology. Two steps are used in this task's identification process. An ideal dataset encompassing all recordings of various bird species was produced in the first phase. The sound clip was then put through several sound pre-processing procedures, including pre-emphasis, framing, silence removal, and reconstruction. For each of the reconstructed sound clips, a spectrogram was produced. In the second step, spectrogram images are given as input to our models. Then the model analyzes the spectrogram and tries to predict the bird species.

KEYWORDS: Bird species identification, Spectrograms, Convolutional Neural Network, Efficient NetB2.

I. INTRODUCTION

Identifying bird species from audio recordings is a challenging field of research. Many researchers prefer to classify birds via audio rather than images because it is not always easy to capture images while their audio is usually detectable. While noise in the audio can be an issue, audio signals provide much more information about a bird, as it can be further classified into songs, calls, and sounds. Having this extra set of classifications also makes identifying birds that much more difficult. The frequency range and volume level within a single bird can be very dynamic. Bird watching and recording is highly popular for amateurs and scientists, especially in matters involving conservation status. Therefore, being able to improve audio classification on birds, would assist researchers in understanding certain species and the environments they live in better.

In this research, we suggested a way to automate the entire process of detecting bird sounds using sound processing, deep learning, and transfer learning models. The initial step is to compile all the recordings into a database. Afterward, acoustic preprocessing methods like pre-emphasis, framing, silence removal, and reconstruction are applied to these recordings. For audio recordings, spectrograms were created, and these spectrograms served as input to our trained model, which used the spectrogram to estimate the species of bird.

II. LITERATURE REVIEW

2.1 Survey of Existing System

- **Convo-Codes: Audio Hashing for Bird Species Classification [1]**

  This paper deals with bird species classification using convo coder audio hashing; the pro- posed framework utilizes archetypal analysis, a matrix factorization technique, to obtain convex-sparse representations of a bird vocalization. This process is divided into the following steps: -

  1. Preprocessing and dictionary learning.
  2. Feature Representation.
  3. Generating convo codes and populating hash tables.
  4. Classification.

- **Automated Bird Species Identification using Audio Signal Processing and Neural Networks [2]**

  In this research, methods for identifying birds have been examined, along with an automatic system for recognizing their species. A two-stage identification approach is used in this work. The first step was creating an ideal dataset that included all the bird species' sound recordings. In the subsequent stage, a neural network was set up and given the spectrograms as input. The Convolutional Neural Network (CNN) categorizes the sound clip and determines the species of bird based on the input features.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |
|| Volume 11, Issue 4, April 2023 ||
| DOI: 10.15680/IJIRCCE.2023.1104114 |

- **Bird Species Recognition Using Unsupervised Modeling of Individual Vocalization Elements [3]**

  This model involves acoustic modeling for the identification of bird species from audio recordings. It consists of HMM (Hidden Markov Model) based clustering and is trained by using a maximum likelihood procedure. Acoustics are decomposed into sinusoidal components. They divided this process into four parts:-

  1. Acoustic scene decomposition.
  2. Feature Representation.
  3. Acoustic modeling and Multiple bird species recognition.

- **Visualization of Audio Records for Automatic Bird Species Identification [4]**

  In this model, a set of bird audio records is processed to extract vector features. A matrix is built with all obtained vectors. The obtained matrix is then used as an input to project audio records in a 2D coordinate system and at the same time to calculate a distance matrix Some species are shown in a map of Colombia.

III. PROPOSED SYSTEM

A. Introduction

One of nature's most magnificent species, birds play a significant role in upholding ecological equilibrium on a global scale. Almost 10,000 different bird species can be found worldwide. Birds are a reliable eco-change or the hazardous environment in the lower level’s indicators. For a complete understanding of the local ecosystem, it is crucial to be aware of the various bird species and the amount of them in a certain location. Nonetheless, we frequently hear birds rather than seeing them. Because of this, practicing bird identification by listening to their calls and noises is a great idea. There have been numerous attempts to identify and categorize birds from their natural habitat, however there are many types of background noise, including environmental, other living objects, and sounds from various bird species. As a result, domain specialists classify the recordings the majority of the time, which is a laborious and time-consuming process. Thus, automatic bird detection is essential for accurate and quick identification.

This project's analysis of the impact of acoustic feature selection in a deep learning-based bird-call categorization challenge is one of its key goals. The goal is to provide readers a thorough understanding of feature selection rather than to reach the highest level of accuracy. Afterwards, the distributed Neural network is given these features as input to classify the bird species. No matter the type of bird cry, location, recording method, or kind of noise used, only bird recordings were used in the study. As a result, the suggested solution uses a CNN-based network to provide a general framework for classifying bird calls. There are four primary steps to the proposed work. The initial phase is the procedure of collecting the bird sound data. After that, pre-processing the data should be done to make the playback more effective. The subsequent steps involve the extraction of features from sound pattern data and the classification of those characteristics using machine learning methods.

B. Architecture

To explain the working of project, our architecture is divided into the following steps:

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104114 |

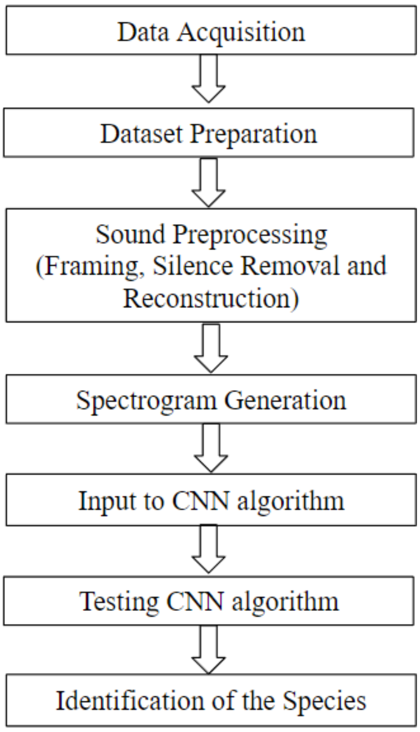


Figure3.1: Flowchart

- **Data Collection:**
  Through a Python wrapper from GitHub [8] we downloaded our dataset from xeno-canto.org, with approximately 600 audio files per bird, totaling 6000 data samples. We chose to limit the number of birds, because we wanted enough samples per bird, but did not have the computing resources to store data for more than 10 birds. Main motive of this step is to collect the data and organize it in such a way that it would help in training our model.

- **Data Preprocessing:**
  There is no method to remove the background noise from a natural recording before using it, thus the audio clips cannot be cleaned up. The original sound recordings also contained undesired sounds. While attempting to hear a wild bird's cry, in particular, it is imperative to remove or reduce background noise. The sound recordings must be pre-processed to ensure reasonable device efficiency. The initial sound recordings contain unwanted noises. While trying to hear a wild bird's cry, it is very important to remove or reduce background noise from the targeted sound. The first step in our data pre-processing was to shorten each sample to about 10 seconds (waveform length of 227,556). Butter, a frequency-based pass band-pass filter, has a highly flat frequency response in the passband and is used to remove undesirable background noise by filtering out a frequency spectrum. The audio signal must be stripped of all other parts but the necessary ones.

- **Feature Extraction:**
  Before bird classification can start, one of the first things that must be done is the type of attribute in particular bird species expansion. Spectrograms, Mel-spectrograms, and MFCC are the key features we employed. The frequencies of a signal are shown visually in a spectrogram over time. Mel spectrograms are created by converting spectrogram frequencies to the Mel scale. They are used to determine the perceived scale of the audio's pitches and the distances between them. Phonemes are represented by MFCC characteristics, which also describe the general sound form. Due to its ability to identify audio similarities, MFCC is frequently employed in speech recognition and information retrieval systems, including genre

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104114 |

categorization. Due to its ability to identify audio similarities, MFCC is frequently employed in speech recognition and information retrieval systems, including genre categorization. Phonemes, or unique sound units, are represented by MFCC characteristics, which also characterize the general contour of the sound. Due to its ability to identify audio similarities, MFCC is frequently employed in speech recognition and information retrieval systems, including genre categorization. Due to a lack of resources (memory), we could only use the first channel of the audio waveforms in our models.

- **Classification Techniques:**

After the process is complete, the various classifications are taken into account to determine which strategy is most popular. The operating principles of the data classifiers used to create and test the training and test sets are briefly explained in these subsections.

CNN: CNN (Convolutional Neural Networks) is a form of deep learning model for processing input that has a grid pattern, such as photographs, which is inspired by the organization of animal visuals and meant to automatically and adaptively learn spatial hierarchies of features, from low- to high-level patterns. Convolution, pooling, and fully linked layers are the three types of layers (or "building blocks") that make up a standard CNN. Convolution and pooling layers in order one and two do feature extraction, whereas a fully connected layer in order three maps the extracted features into the output, such as classification. In CNN, which is made up of a stack of mathematical operations, including convolution, a specialized kind of linear operation, a convolution layer is crucial. Since a feature may appear anywhere in a digital image, the pixel values are stored in a two-dimensional (2D) grid, or array of numbers, and a small grid of parameters called the kernel, an optimizable feature extractor, is applied at each image position, making CNNs extremely effective for image processing. Extracted features may gradually and hierarchically become more sophisticated as one layer feeds its output into the following layer. Training is the process of minimizing the difference between outputs and ground truth labels using an optimization technique like backpropagation and gradient descent, among others. It involves improving parameters such as kernels.

EfficientNet B2 Model: EfficientNet is a convolutional neural network design and scaling technique that uses a compound coefficient to consistently scale all depth, breadth, and resolution dimensions. The EfficientNet scaling method evenly scales network breadth, depth, and resolution using a set of preset scaling coefficients, in contrast to standard practice, which scales these variables arbitrarily. For example, if we want to use 2N times more computational resources, then we can simply increase the network depth by αN, width by βN, and image size by γN, where α,β,γ are constant coefficients determined by a small grid search on the original small model. EfficientNet uses a compound coefficient ϕ to uniformly scales network width, depth, and resolution in a principled way. The reasoning behind the compound scaling method is based on the supposition that larger input images necessitate more channels and layers in order to catch more fine-grained patterns on the larger image. The sole architectural distinction between it and the EfficientB2 model is the variation in the number of feature maps (channels), which raises the number of parameters.

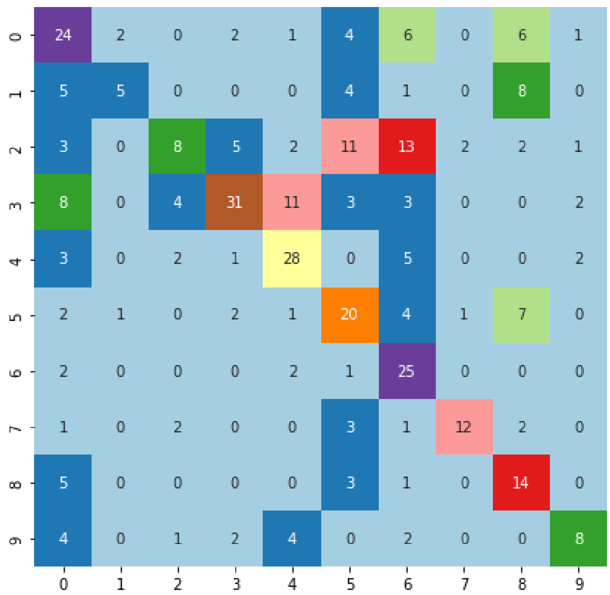
IV. EXPERIMENTS AND RESULTS

Due to the fact that we have two models, we must assess which model performs best, hence we are utilizing the confusion matrix to do so.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

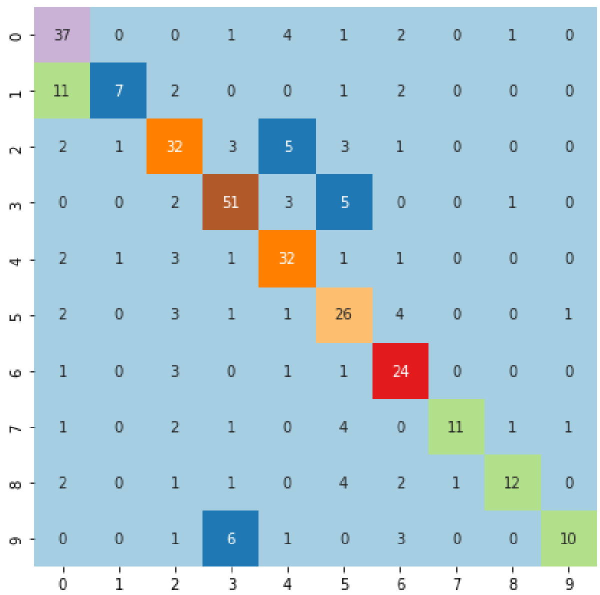
|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104114 |



CNN

Fig 4.1 Confusion matrix (CNN)



EfficientNet B2

Fig 4.2 Confusion matrix (EfficientNet B2)

The right predictions in the EfficientNet B2 model outnumber the right predictions in the CNN model, which is why it performs better than the CNN model, according to the confusion matrices mentioned above. As a result, we opted for the EfficientNet B2 model rather than the CNN model.

Snapshots of our project :

- This is the home screen of our project. Here we have to upload the audio file and then click on "Discover".

- A spectrogram is created using the audio file after it has been uploaded. The spectrogram signals are used to identify the species of bird. The created spectrogram, the bird's name and details, as well as the prediction's accuracy, are all displayed after we upload the audio file to the system.

Fig 4.3: Uploading the Audio file to our site

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104114 |

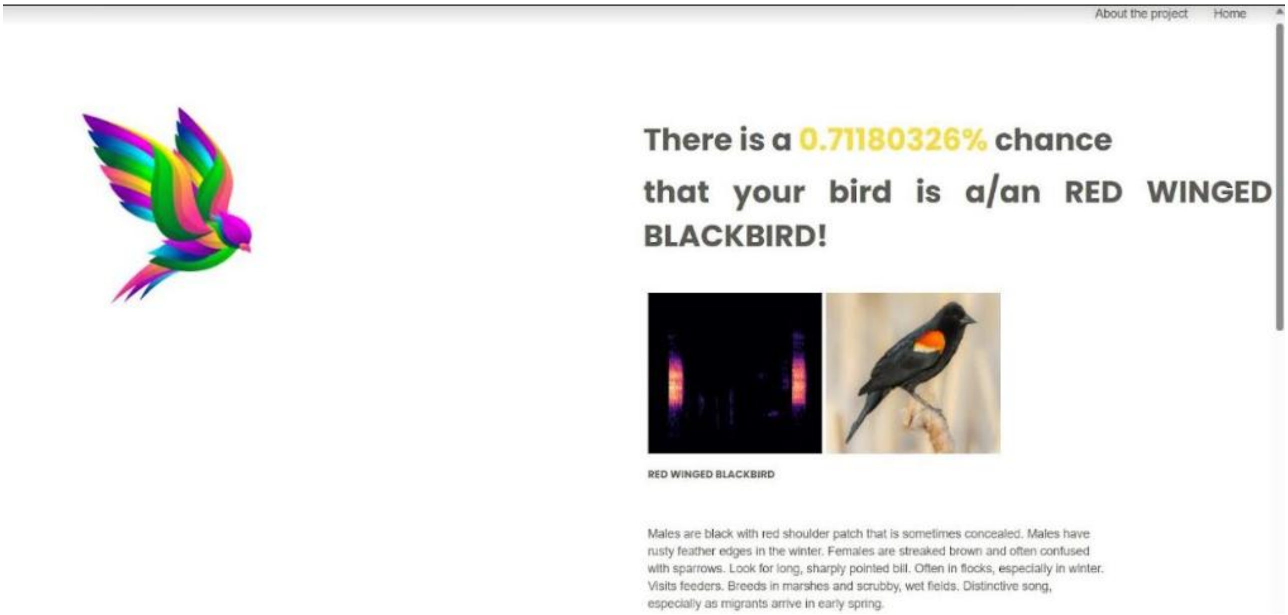


Fig 4.4: Final Output

Algorithm:

- **Step 1:** Audio file will be taken as an input from the user.
- **Step 2:** After the audio is uploaded, audio is converted into segments of equal duration (ap- prox 5 to 10 sec)
- **Step 3:** From the audio segments generated in Step 2, MFCC Features are extracted.
- **Step 4:** Using the MFCC Features extracted in the above step, a spectrogram is generated. This spectrogram is given as an input to our model.
- **Step 5:** The model then tries to extract maximum features from the spectrogram with the help of various layers like convolutional layer, ReLU layer, Pooling layer etc.
- **Step 6:** The model attempts to estimate the bird species using these extracted features and displays the result with prediction accuracy.

V. CONCLUSION

Overall, we attempted to accurately categorize birds using only audio data from xeno-canto by adopting Classification model and modifying CNN models from prior work. This allowed us to use a variety of models, including CNN Classification and Transfer Learning Models, with various degrees of effectiveness. We talked about our earlier study and the pretreatment approach for our particular dataset before moving on to the modeling stage.

REFERENCES

1. Charbuty B, Abdulazeez A. “Classification Based on Decision Tree Algorithm for Machine Learning,” J. Appl. Sci. Technol. Trends. 2021;01:20–28, 2021.
2. Saleem SI, Abdulazeez AM, Orman Z. “A New Segmentation Framework for Arabic Handwritten Text Using Machine Learning Techniques; 2021.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104114 |

3. Maulud D, Abdulazeez AM. “A Review on Linear Regression Comprehensive in Machine Learning,” J. Appl. Sci. Technol. Trends. 2020;1(4):140–147.
4. Trappey AJC, Trappey CV, Wu JL, Wang JWC. “Intelligent compilation of patent summaries using machine learning and natural language processing techniques,” Adv. Eng. Informatics. 2019-
5. Abdulkareem NM, Abdulazeez AM. “Machine Learning Classification Based on Random Forest Algorithm : A Review. 2021;128–142.
6. Mahmood MR, Abdulazeez AM, Orman Z. “A New Hand Gesture Recognition System Using Artificial Neural Network,” 2nd Int. Sci. Conf. ZAKHO. 2017;1–14.
7. Lasseck M. “Audio-based bird species identification with deep convolutional neural networks,” CEUR Workshop Proc. 2018;2125.
8. Zeebaree DQ, Haron H, Abdulazeez AM, Zebari DA. “Machine learning and Region Growing for Breast Cancer Segmentation,” 2019 Int. Conf. Adv. Sci. Eng. ICOASE 2019;88–93.
9. Ibrahim I, Abdulazeez A. “The Role of Machine Learning Algorithms for Diagnosing Diseases,” J. Appl. Sci. Technol. Trends. 2021;2(01):10–19.
10. Piczak KJ. “Recognizing bird species in audio recordings using deep convolutional neural networks,” CEUR Workshop Proc. 2016;1609:534–543.
11. Kahl S, Wilhelm-Stein T, Klinck H, Kowerko D, Eibl M. Recognizing birds from sound - The 2018 BirdCLEF baseline system,” arXiv; 2018.
12. Abdulqader DM, Abdulazeez AM, Zeebaree DQ. “Machine learning supervised algorithms of gene selection: A review,” Technol. Reports Kansai Univ. 2020;62(3);233–244.
13. Sprengel E, Jaggi M, Kilcher Y, Hofmann
14. T. “Audio based bird species identification using deep learning techniques,” CEUR Workshop Proc. 2016;1609:547–559.
15. Salim NOM, Abdulazeez AM. “Human Diseases Detection Based On Machine Learning Algorithms : A Review. 2021;102–113.