UNIVERSITY INSTITUTE OF TECHNOLOGY

The University of Burdwan

DEPARTMENT OF INFORMATION TECHNOLOGY

2020-2024 BATCH

Project Report On

Bird Image Retrieval and Recognition Using Deep Learning

A Project under the Guidance of

Prof. (Dr.) Shiladitya Pujari

Assistant Professor,

DEPARTMENT OF INFORMATION TECHNOLOGY
UNIVERSITY INSTITUTE OF TECHNOLOGY
THE UNIVERSITY OF BURDWAN GOLAPIAG (NORTH)
BURDWAN 713104, WEST BENGAL

Submitted by

Vivek Kumar [20203003]

UNIVERSITY INSTITUTE OF TECHNOLOGY
The University of Burdwan

DEPARTMENT OF INFORMATION TECHNOLOGY
2020-2024 BATCH

A Project under the Guidance of

Prof. (Dr.) Shiladitya Pujari

Assistant Professor,

DEPARTMENT OF INFORMATION TECHNOLOGY
UNIVERSITY INSTITUTE OF TECHNOLOGY
THE UNIVERSITY OF BURDWAN GOLAPBAG (NORTH)
BURDWAN 713104, WEST BENGAL

"Bird Image Retrieval and Recognition Using Deep Learning"

Submitted by

Vivek Kumar [20203003]

University Institute of Technology , University of Burdwan

DEPARTMENT OF INFORMATION TECHNOLOGY

BACHELOR OF ENGINEERING (B.E.)

CERTIFICATE OF APPROVAL

This is to certify that the project entitled " **Bird Image Retrieval and Recognition Using Deep Learning** " submitted by **Vivek Kumar [20203003]** under the guidance and supervision **Prof. (Dr.) Shiladitya Pujari** as partial fulfilment for the award of the degree of **BACHELOR OF ENGINEERING** in **INFORMATION TECHNOLOGY at UNIVERSITY INSTITUTE OF TECHNOLOGY**. The University of Burdwan is a record of the work of the students which has been carried out under any supervision.

I hereby forward this project.

Date:…………..

Place :………….

(Project Guide)
Prof. (Dr.) Shiladitya Pujari

Assistant Professor,

DEPARTMENT OF INFORMATION TECHNOLOGY
UNIVERSITY INSTITUTE OF TECHNOLOGY

University Institute of Technology , University of Burdwan

CERTIFICATE

This is to certify that the work embodied in the final year project " **Bird Image Retrieval and Recognition Using Deep Learning** ", submitted by **Vivek Kumar** to the Department of Information Technology are carried out under my direct Supervision and guidance. The project work has been prepared as per regulation of The University of Burdwan and I strongly recommend that this project work can be accepted in partial fulfillment.

Date.............

Date................

Head of the Department :
Mrs. Kasturi Ghosh
Professor & In-charge, Dept. of IT
University Institute of Technology

Project Guider :
Prof. (Dr.) Shiladitya Pujari
Lecturer, Dept. of IT
University Institute of Technology

ACKNOWLEDGEMENT

We express our pleasure and gratitude to submit our project " **Bird Image Retrieval and Recognition Using Deep Learning** ". We are extremely obliged to **Prof. (Dr.) Shiladitya Pujari** sir who have given us the opportunity to work on this project. They have been a perpetual source of inspiration in taking this project to such great heights on behalf of the institution UNIVERSITY INSTITUTE OF TECHNOLOGY, BURDWAN UNIVERSITY. Finally, we would like to thank THE UNIVERSITY OF BURDWAN for giving us this subject as a paper as it has not only given us an experience to work in a project but also will help in our near future.

Date…………….

Vivek Kumar

Table of Contents

DECLARATION

We clarify that:

1. The work contained in the thesis is original and has been done by ourselves under the general supervision of our supervisor(s).

2. The work has not been submitted to any other Institute for any degree or diploma.

3. We have followed the guidelines provided by the Institute inwriting the thesis.

4. We have conformed to the norms and guidelines given in theoretical code of conduct of the Institute.

5. Whenever we have used materials (data, theoretical analysis, and text) from other sources, we have given due credit to them by citing them in the text of the thesis and giving their details in the references.

6. Whenever we have quoted written materials from other sources, we have put them under quotation marks and guidance credit to the source by citing them and giving required details in the references.

Abstract

Bird image retrieval and recognition play vital roles in various fields such as wildlife conservation, ecological research, and birdwatching .This project focuses on the development of a bird image retrieval and recognition system employing deep learning techniques. The methodology involves constructing a comprehensive dataset of annotated bird images to facilitate model training. A pre-trained convolutional neural network (CNN) is utilized for feature extraction, followed by fine-tuning on the bird dataset to enhance recognition accuracy.

The trained model serves as the core component for recognizing various bird species. To enable efficient image retrieval, techniques such as transfer learning and image embeddings are employed, enhancing similarity scoring mechanisms. Implementation is carried out using popular deep learning frameworks such as TensorFlow or PyTorch, with libraries like Keras providing additional support.

This project aims to contribute to the field of avian image analysis by providing an effective solution for bird image retrieval and recognition through the utilization of deep learning methodologies.

objective

The objective of implementing bird image retrieval and recognition using deep learning is to develop a system that can accurately identify and categorize bird species based on visual information. This involves leveraging deep neural networks, particularly convolutional neural networks (CNNs), to extract meaningful features from bird images. The primary goals include: **Species Identification**: Design a model capable of recognizing various bird species from images, enabling automated identification and classification.

Accuracy Improvement: Enhance the accuracy of bird species recognition by leveraging deep learning techniques, minimizing misclassifications, and improving generalization across diverse datasets.

Feature Extraction: Implement advanced feature extraction using CNNs to capture intricate patterns and characteristics present in bird images, facilitating robust recognition across different poses, lighting conditions, and backgrounds.

Transfer Learning: Incorporate transfer learning techniques to leverage pre-trained models and optimize performance, reducing the need for extensive labeled data and computational resources.

Real-world Applicability: Develop a system that is practical for real-world scenarios, considering variations in environmental conditions, diverse bird species, and potential challenges in image quality.

User-Friendly Retrieval: Create an intuitive retrieval system that allows users to efficiently search for bird species through images, promoting accessibility and usability.

Dataset Diversity: Ensure the inclusion of a diverse and representative dataset to train the model, addressing potential biases and improving its ability to handle a wide range of bird species.

Contribution to Conservation: Contribute to biodiversity research and conservation efforts by providing a tool that aids in the identification and monitoring of bird species, supporting ecological studies and conservation initiatives. By achieving these objectives, the bird image retrieval and recognition system can become a valuable tool for ornithologists, researchers, and conservationists, aiding in the understanding and preservation of avian biodiversity .

Motivation

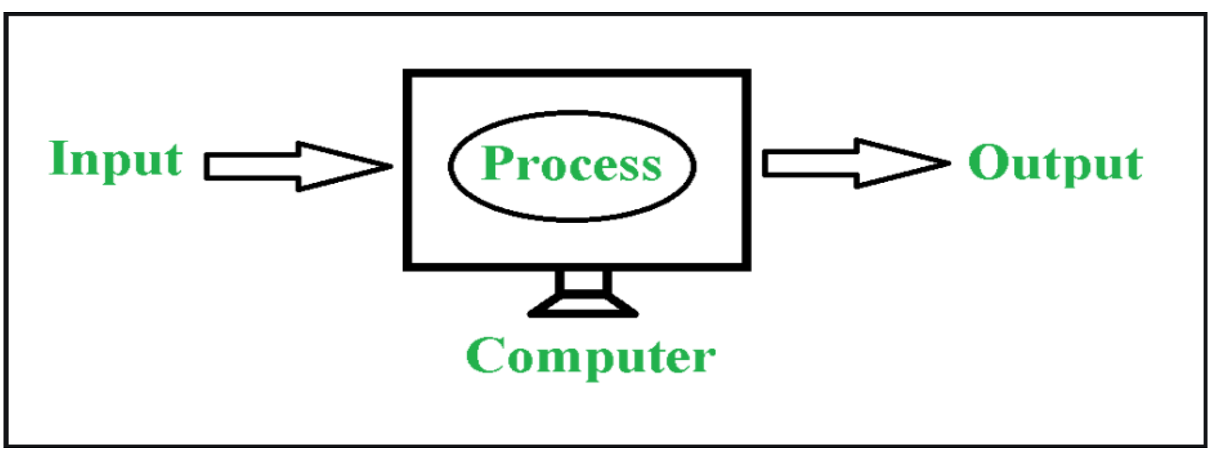
The motivation for employing deep learning in bird image retrieval and recognition arises from the unique advantages offered by deep neural networks. Deep learning models, particularly convolutional neural networks (CNNs), excel in automatically extracting complex features from images, a crucial aspect when dealing with the diverse shapes, colors, and patterns exhibited by birds. The hierarchical nature of deep learning architectures, with multiple layers learning progressively abstract representations, aligns well with the hierarchical features found in bird images, enabling the model to discern intricate patterns at different levels of complexity. Moreover, the end-to-end learning capability eliminates the need for manual feature engineering, making the system adaptable to various bird species and image variations. Deep learning's scalability is essential when dealing with extensive bird image datasets, facilitating the training of models on diverse species and variations, enhancing generalization. The transfer learning paradigm, where pre-trained models on large image datasets are fine-tuned for bird recognition tasks, leverages knowledge gained from general image recognition, especially beneficial when labeled bird datasets are limited. Additionally, deep learning models showcase adaptability to varied environmental conditions, crucial for robust performance in real-world scenarios where birds inhabit diverse habitats. Finally, the state-of-the-art performance demonstrated by deep learning models, particularly CNNs, in image recognition tasks further underscores their suitability for achieving high accuracy in identifying and classifying bird species.

What is Image Classification?

• Image classification is a machine learning technique that assigns a set of predefined categories to open-ended images. Image classifiers can be used to organize, structure, and categorize pretty much any kind of image – from photographs, medical scans and diagrams, to visuals found all over the web.

• Image classification is one of the fundamental tasks in computer vision with broad applications such as identifying the species of animals, recognizing objects in a scene, determining the location of landmarks, and describing visual content, among others.

• An image classifier can take an image as input, analyze its content, and then automatically assign relevant tags, such as User Interface (UI) and Easy To Use, based on the visual features and patterns detected in the image.



SURVEY REPORTS

1. A Review on Image Dataset to analyze birds using different images from different angles for same bird in Birds Datasets.

Processing **by**

Vivek kumar, Jatan Tiwari, Devanshi Ananad, Suvrajit Ghosh, Soumyadip Som, Riju Nandi

- **Purpose:** To validate the usefulness of the convolutional neural network (CNN) to classify the birds' images using the images taken by users
- **Methods:** design: comparison of convolutional neural network (CNN).
- **Dataset Used:** Training dataset include 5K species of birds' name along with their images. All data extracted from website ebirds by using "Web scraping" methodology.
  - **Observation procedure:** The bird identification software utilizing Convolutional Neural Networks (CNN) follows a systematic procedure. Commencing with the collection of a diverse, labeled dataset, the images undergo preprocessing involving resizing, normalization, and augmentation. The CNN model architecture, tailored for bird image classification, includes convolutional and pooling layers, along with fully connected layers for multi-class classification. Training the model involves splitting the dataset and monitoring validation metrics to prevent overfitting. Post-training evaluation assesses the model's generalization performance using metrics like accuracy, precision, recall, and F1 score, while user-friendly interface development allows users to upload or capture images for identification. Encouraging user

feedback establishes a continuous improvement loop, and performance optimization ensures real-time efficiency. Regular updates, maintenance, and thorough documentation complete the comprehensive procedure, fostering a reliable and effective bird identification software.

- **Conclusion:** This study validates a CNN for bird image classification using a 5K-species training dataset from eBirds. The observation procedure involves preprocessing, training, and user-friendly interface, ensuring reliable and efficient bird identification software.

Why is Image Classification Important?

Approximately 80% of information is unstructured, with images being a prevalent form of unstructured data. Analyzing, understanding, and organizing image data manually is challenging and time-intensive, limiting its utilization by companies.

- **Why use deep learning for image classification?**
- **Scalability:** Manual image analysis is slow and less accurate. Deep learning can efficiently process millions of images at a fraction of the cost, offering scalability to businesses of any size
- **Real-time analysis**Deep learning enables real-time monitoring, crucial for promptly identifying and responding to critical situations, such as bird species identification, ensuring swift actions based on immediate insights.
- **Consistent criteria**
  Human annotators may introduce errors and inconsistency due to distractions and subjectivity. Deep learning applies consistent criteria to

image data, ensuring accuracy and reliability once the model is properly trained

How Does Image Classification Work?

• Image classification can be executed through manual or automatic methods.
• Manual image classification involves a human annotator interpreting image content and categorizing it accordingly, which, although effective, is time-consuming and costly.
• Automatic image classification utilizes deep learning, computer vision, and other AI-driven techniques for faster, more cost-effective, and accurate classification.
• Three types of systems are prevalent in automatic image classification:
Rule-based systems: These systems use handcrafted rules to organize images into predefined categories, relying on semantically relevant features to identify content-based categories.

Deep Learning-based systems: Instead of manual rules, deep learning image classification learns from labeled examples, discerning associations between visual features and predefined categories. Deep neural networks, such as Convolutional Neural Networks (CNNs), excel in learning hierarchical representations from image data.

Hybrid systems: Combining rule-based and deep learning approaches, hybrid systems leverage both predefined rules and learned features for image categorization, offering a balance between interpretability and accuracy.

Deep Learning Image Classification Algorithm

Convolutional Neural Networks (CNNs) are a class of deep learning models specifically designed for image classification tasks. CNNs have proven highly effective in recognizing patterns and extracting hierarchical features from images. Here's a breakdown of key concepts in CNNs for image classification:

Convolutional Layers: CNNs use convolutional layers to scan input images with filters (also known as kernels) to detect patterns like edges, textures, and more complex structures.

These filters slide over the input image, and the convolution operation computes the dot product between the filter and local regions of the image.

Activation Layers: Activation layers introduce non-linearities to the model, allowing it to learn complex relationships in the data. Common activation functions include ReLU (Rectified Linear Unit).

Pooling Layers: Pooling layers reduce the spatial dimensions of the input by down-sampling, helping in retaining important features and reducing computational complexity.

Max pooling is a common technique, where the maximum value in a region is retained, discarding less important information.

Fully Connected Layers: After multiple convolutional and pooling layers, the high-level reasoning in the neural network is captured in fully connected layers.

These layers connect every neuron to every neuron in the previous and subsequent layers, creating a high-level representation of the input.

Flattening: Before passing data to fully connected layers, the output from previous layers is flattened into a vector.

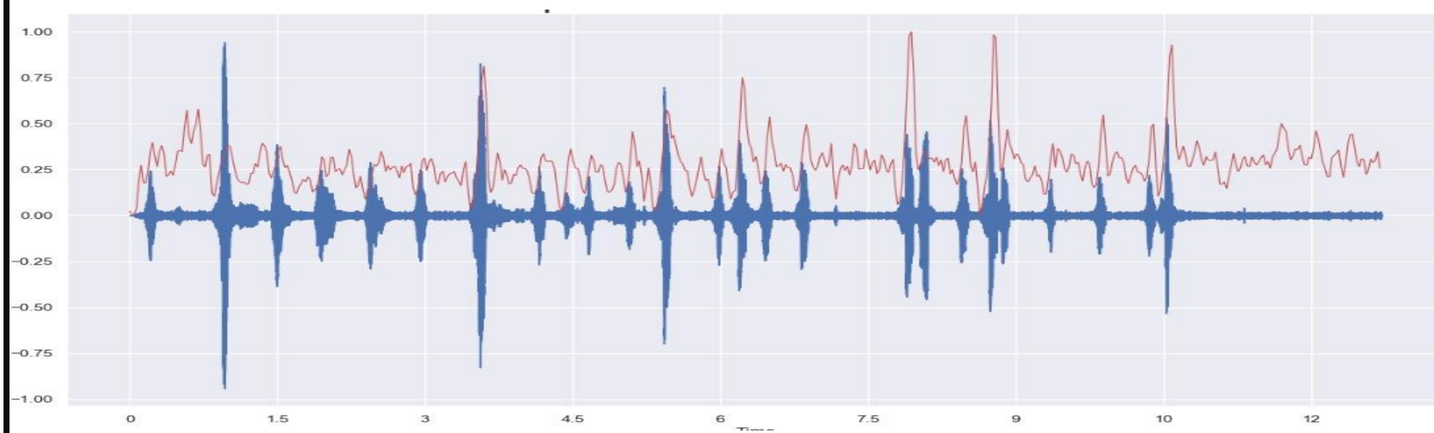
Softmax Layer: In the final layer, a softmax activation function is often used for multi-class classification problems. It converts the network's output into probability scores for each class.

Loss Function: The model is trained by minimizing a loss function, such as categorical cross-entropy, which measures the difference between predicted probabilities and actual class labels.

Training: During training, the model adjusts its internal parameters (weights and biases) based on the backpropagation algorithm and optimization techniques (e.g., gradient descent) to minimize the loss function.

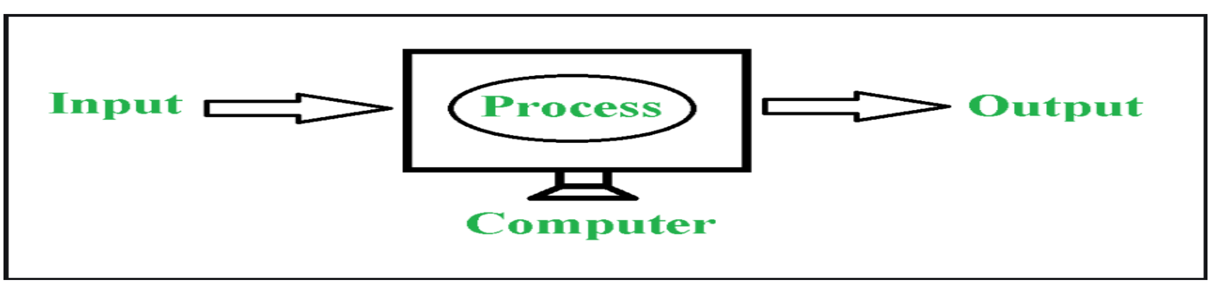
CNNs excel in capturing hierarchical and spatial features in images, making them well-suited for image classification tasks. They have been pivotal in achieving state-of-the-art results in various computer vision applications, including object recognition, image segmentation, and more. Common architectures like AlexNet, VGGNet, GoogLeNet, and ResNet have further advanced the field of image classification using CNNs.

<u>What is Sound Classification?</u>



Sound classification is a machine learning technique that assigns a set of predefined categories to audio data. Sound classifiers can be employed to organize, structure, and categorize various types of sound – from music recordings, environmental sounds, and spoken words. Sound classification is one of the fundamental tasks in audio processing with broad applications, such as identifying musical genres, detecting environmental noises, and transcribing spoken language. It plays a crucial role in tasks like classifying the genre of a song, identifying the speaker in a recording, or distinguishing between different types of environmental sounds.

A sound classifier can take an audio snippet as input, analyze its acoustic features, and automatically assign relevant tags, such as Music, Speech, or Environmental Sounds. This technology finds applications in diverse fields, including speech recognition, music recommendation systems, and surveillance systems that can identify specific sounds in an environment.



SURVEY REPORTS

A Review on Sound Dataset to analyze birds using sounds of the same bird in Birds Datasets.

Processing **by**

Vivek Kumar, Jatan Tiwari, Devanshi Ananad, Suvrajit Ghosh, Soumyadip Som, Riju Nandi

- **Purpose:** To validate the usefulness of TensorFlow to classify the birds' sounds using the sounds taken by users
- **Methods:** design: comparison of TensorFlow.
- **Dataset Used:** Training dataset include 5K species of birds' name along with their Sounds. All data extracted from website ebirds by using "Web scraping" methodology.
    - **Observation procedure:** The sound identification software, leveraging Fourier Transform-based Neural Networks (FTNN) and TensorFlow, follows a systematic approach. Starting with the assembly of a diverse, labeled dataset, audio files undergo preprocessing involving the application of the Fourier Transform for spectral analysis, normalization, and augmentation. The FTNN model architecture, tailored for sound classification, incorporates layers designed for Fourier feature extraction and temporal pooling, coupled with fully connected layers for multi-class classification. Model training involves dataset partitioning, and the training process is monitored using validation metrics to prevent overfitting. Post-training evaluation assesses the model's generalization performance using metrics like accuracy, precision, recall, and F1

score. The development of a user-friendly interface, implemented with TensorFlow, enables users to upload or capture audio clips for identification. Soliciting user feedback establishes a continuous improvement loop, and performance optimization using TensorFlow ensures real-time efficiency. Regular updates, maintenance, and thorough documentation complete the comprehensive procedure, fostering a reliable and effective sound identification software.

Conclusion: This study validates an FTNN for sound classification using a diverse training dataset sourced from various audio sources, with TensorFlow playing a pivotal role. The observation procedure encompasses Fourier Transform-based preprocessing, training, and a user-friendly interface developed using TensorFlow, ensuring the development of a reliable and efficient sound identification software.

Why is Sound Classification Important?

- Approximately 80% of information is unstructured, with audio data being a prevalent form of unstructured information. Analyzing, understanding, and organizing sound data manually is challenging and time-intensive, limiting its utilization by companies.

Why use deep learning for sound classification?

- **Scalability:** Manual sound analysis is slow and less accurate. Deep learning can efficiently process millions of audio clips at a fraction of the cost, offering scalability to businesses of any size.
- **Real-time analysis:** Deep learning enables real-time monitoring, crucial for promptly identifying and responding to critical situations, such as

detecting specific sounds in surveillance or emergency scenarios. This ensures swift actions based on immediate insights.

- **Consistent criteria:** Human annotators may introduce errors and inconsistency due to distractions and subjectivity when classifying sounds. Deep learning applies consistent criteria to audio data, ensuring accuracy and reliability once the model is properly trained. This is particularly valuable for tasks such as identifying speech patterns or distinguishing between various environmental sounds.

How Does Sound Classification Work?

• Sound classification can be carried out through manual or automatic methods.

• Manual sound classification involves a human annotator interpreting audio content and categorizing it accordingly, which, while effective, is time-consuming and costly.

• Automatic sound classification utilizes deep learning, audio processing, and other AI-driven techniques for faster, more cost-effective, and accurate classification.

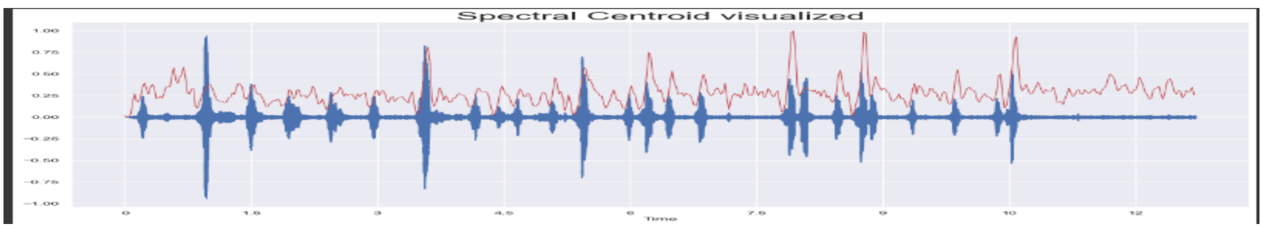
• Three types of systems are prevalent in automatic sound classification:
Rule-based systems: these systems use handcrafted rules to organize audio clips into predefined categories, relying on semantically relevant features to identify content-based categories.

Deep Learning-based systems: Instead of manual rules, deep learning sound classification learns from labeled examples, discerning associations between acoustic features and predefined categories. Deep neural networks, such as Convolutional Neural Networks (CNNs), excel in learning hierarchical representations from audio data.

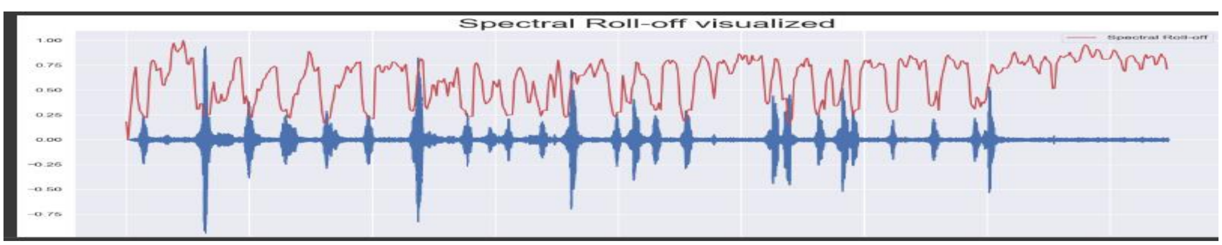
Hybrid systems: Combining rule-based and deep learning approaches, hybrid systems leverage both predefined rules and learned features for sound categorization, offering a balance between interpretability and accuracy in identifying various sound patterns.

Deep Learning Sound Classification Algorithms

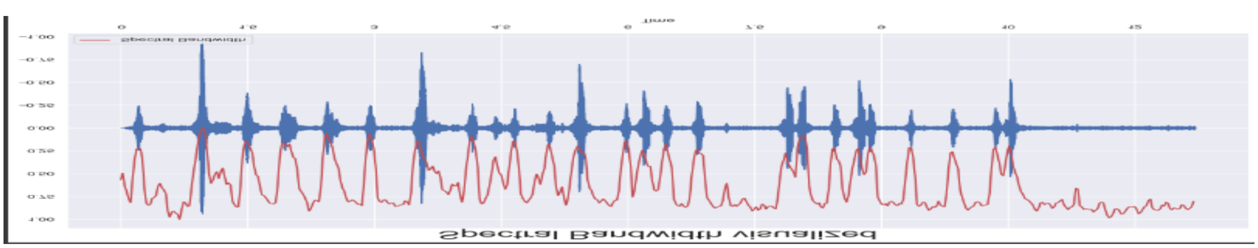
Spectral Centroid: The spectral centroid represents the center of mass of the spectrum and provides information about the "brightness" or "color" of a sound.



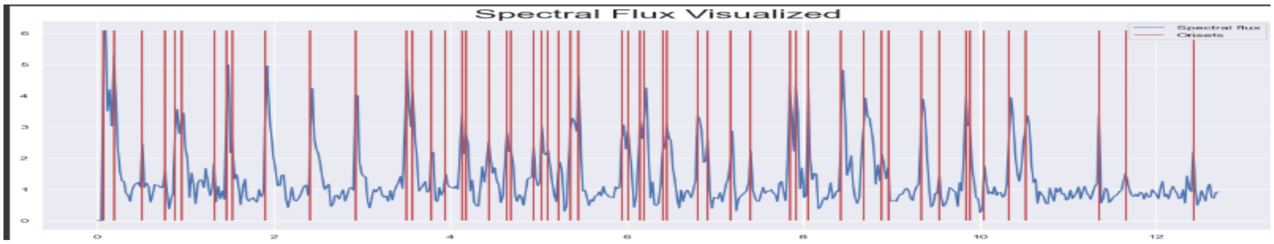
Spectral Roll-off: Spectral roll-off measures the frequency below which a certain percentage of the total spectral energy is contained, characterizing the "edge" or "sharpness" of the spectral content.



Spectral Bandwidth: Spectral bandwidth quantifies the width of the frequency range in a signal, indicating the breadth of the spectral distribution.



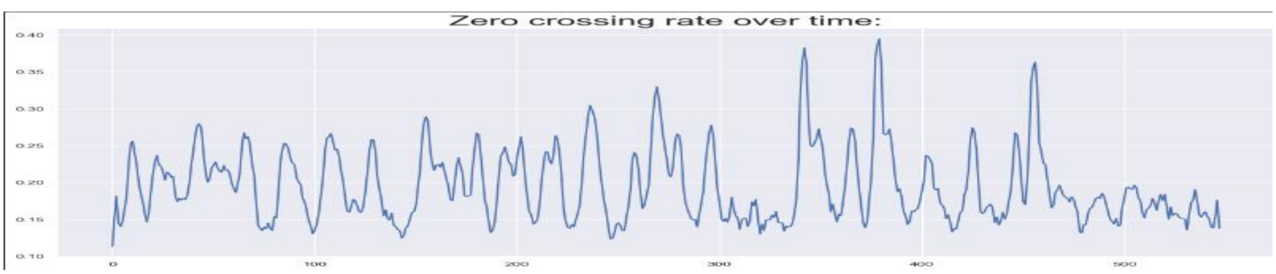
Spectral Flux: Spectral flux measures the change in spectral content between consecutive frames of an audio signal, indicating dynamic changes and onset of new events.



Spectral Contrast: Spectral contrast measures the difference in amplitude between peaks and valleys in the spectrum, capturing perceptually relevant tonal characteristics.

Spectral Flatness: Spectral flatness quantifies how flat or peaked the spectrum of an audio signal is, indicating a constant or variable energy distribution.

Zero-crossing Rate: Zero-crossing rate calculates the rate at which the audio signal changes its sign, useful for distinguishing between voiced and unvoiced sounds.



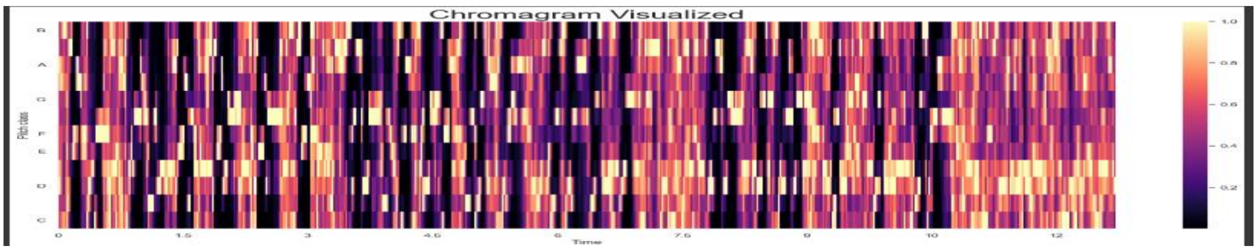
Autocorrelation: Autocorrelation measures the similarity between an audio signal and a delayed version of itself, capturing periodicities in the signal.

Fundamental Frequency: The fundamental frequency represents the lowest frequency component in a sound, crucial for tasks like pitch estimation.

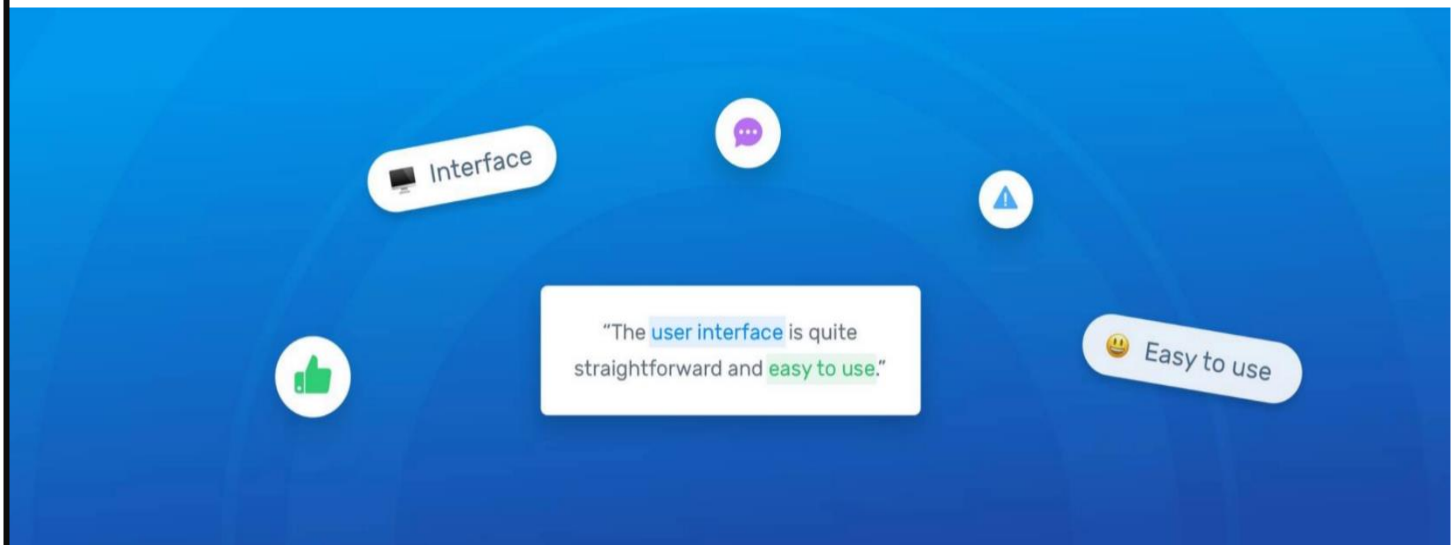
Tempo Estimation: Tempo estimation identifies the beats per minute (BPM) of a musical piece, characterizing the rhythmic structure.

Mel-Frequency Cepstral Coefficients (MFCCs): MFCCs capture the spectral characteristics of a signal, emphasizing perceptually relevant features for speech and audio processing.

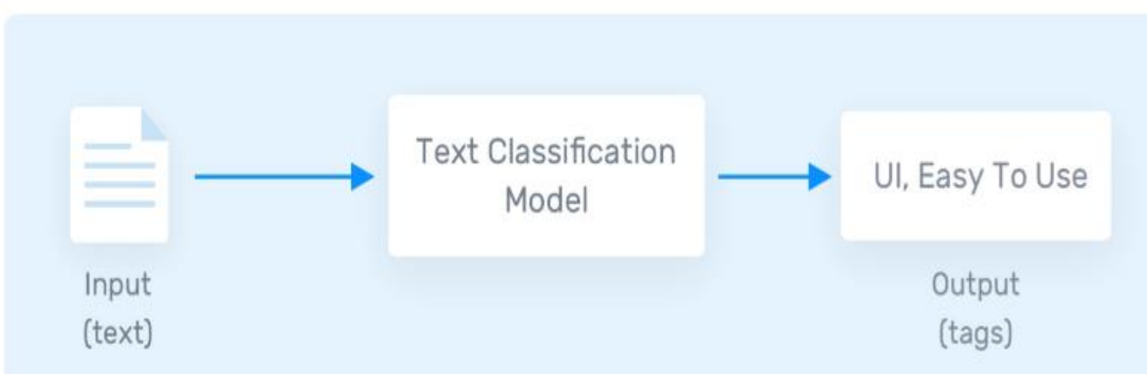
Chromogram: Chromograms represent the energy distribution of different pitch classes in an audio signal, useful for music analysis tasks like chord recognition and key estimation.



What is Text Classification ?



- Text classification is a machine learning technique that assigns a set of predefined categories to open-ended text. Text classifiers can be used to organize, structure, and categorize pretty much any kind of text – from documents, medical studies and files, and all over the web .
- Text classification is one of the fundamental tasks in natural language processing with broad applications such as finding the birds scientific name, species name, location of birds and description of birds etc.
- A text classifier can take this phrase as an input, analyse its content, and then automatically assign relevant tags, such as *UI* and *Easy To Use*.



SURVEY REPORTS

1. A Review on Text Classification from text of Scientific name of Birds species using Birds Datasets.

Processing **by**

Vivek kumar, Jatan Tiwari, Devanshi Ananad, Suvrajit Ghosh, Soumyadip Som , Riju Nandi

- **Purpose :** To validate the usefulness of the "Naive Bayes" classifier to classify the text data and extract some valuable information related to Birds .
- **Methods:** design: comparison of "Naive Bayes" algorithms.
- **Dataset Used:** Training dataset include 5K species of birds name but scientific name of all birds species are different and testing dataset includes 30 % of training dataset. All data extracted from website ebirds by using "Webscraping" .
- **Observation procedure:** Using the training dataset, classifiers were built to discriminate between the scientific name of birds. Like two birds species names are same but the scientific name of both birds are unique. . Multinomial Naïve Bayes or Support vector machine(SVM) algorithms are used for classify the scientific name of same bird species . Results: with the testing data, the are good formatted and correct scientific name is generated according to birds image .
- **Conclusion :** It is useful to analyze the correct scientific name of same birds species. If we take any two same name birds species . using this

classifier we can distinguish between both the birds and find correct scientific name of birds .

- **Accuracy :** 93%

Why is Text Classification Important?

- It's estimated that around 80% of all information is unstructured, with text being one of the most common types of Unstructure data. Because of the messy nature of text, analyzing, understanding, organizing, and sorting through text data is hard and time-consuming, so most companies fail to use it to its full potential.

Why use machine learning text classification ?

- **Scalability**

  Manually analyzing and organizing is slow and much less accurate.. Machine learning can automatically analyze millions of surveys, comments, names, etc., at a fraction of the cost, often in just a few minutes. Text classification tools are scalable to any business needs, large or small.
- **Real-time analysis**

  There are critical situations that project need to identify as soon as possible and take immediate action . Machine learning text classification can follow your birds species name constantly and in real time, so you'll identify critical information and be able to take action right away.
- **Consistent criteria**

Human annotators make mistakes when classifying text data due to distractions, fatigue, and boredom, and human subjectivity creates inconsistent criteria. Machine learning, on the other hand, applies the same lens and criteria to all data and results. Once a text classification model is properly trained it performs with unsurpassed accuracy.

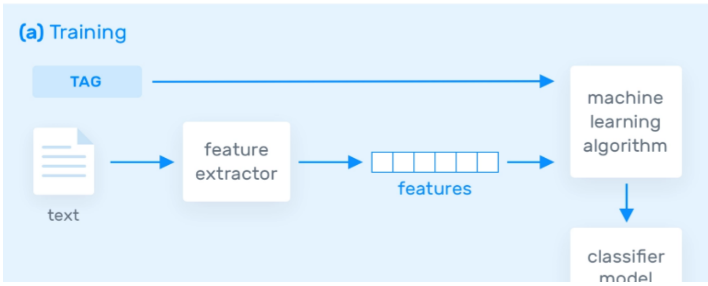
How Does Text Classification Work?

- You can perform text classification in two ways: manual or automatic.
- Manual text classification involves a human annotator, who interprets the content of text and categorizes it accordingly. This method can deliver good results but it's time-consuming and expensive.
- Automatic text classification applies machine learning, natural language processing (NLP), and other AI-guided techniques to automatically classify text in a faster, more cost-effective, and more accurate manner.
- There are many approaches to automatic text classification, but they all fall under three types of systems:
    - Rule-based systems
    - Machine learning-based systems
    - Hybrid systems

1. Rule-based systems

   Rule-based approaches classify text into organized groups by using a set of handcrafted linguistic rules. These rules instruct the system to use semantically relevant elements of a text to identify relevant categories based on its content.

2. Machine learning based systems: Instead of relying on manually crafted rules, machine learning text classification learns to make classifications based on past observations. By using pre-labeled examples as training data, machine learning algorithms can learn the different associations between pieces of text, and that a particular output (i.e., tags) is expected for a particular input (i.e., text). A "tag" is the pre-determined classification or category that any given text could fall into.



Training the Feature . prediction

Machine Learning Text Classification Algorithms

- Some of the most popular text classification algorithms include the Naive Bayes family of algorithms, support vector machines (SVM).

1) Naïve Bayes :

- The Naive Bayes family of statistical algorithms are some of the most used algorithms in text classification and text analysis, overall.
- One of the members of that family is Multinomial Naive Bayes (MNB) with a huge advantage, that you can get really good results even when your dataset isn't very large (~ a couple of thousand tagged samples) and computational resources are scarce.
- Naive Bayes is based on Bayes's Theorem, which helps us compute the conditional probabilities of the occurrence of two events, based on the probabilities of the occurrence of each individual event. So we're calculating the probability of each tag for a given text, and then outputting the tag with the highest probability.

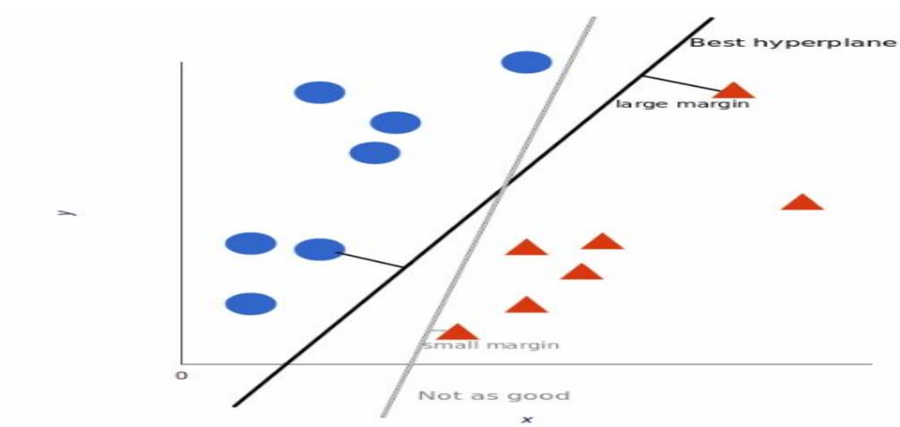
$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

- The probability of A, if B is true, is equal to the probability of B, if A is true, times the probability of A being true, divided by the probability of B being true.

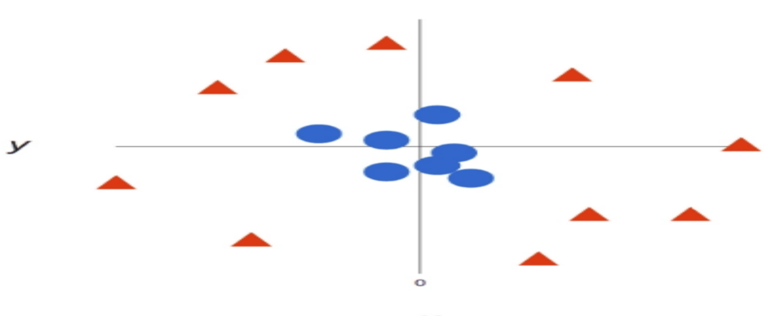
2. **Support Vector Machines**

- Support Vector Machines (SVM) is another powerful text classification machine learning algorithm, because like Naive Bayes, SVM doesn't need much training data to start providing accurate results. SVM does, however, require more computational resources than Naive Bayes, but the results are even faster and more accurate.
- In short, SVM draws a line or "hyperplane" that divides a space into two subspaces. One subspace contains vectors (tags) that belong to a group, and another subspace contains vectors that do not belong to that group.

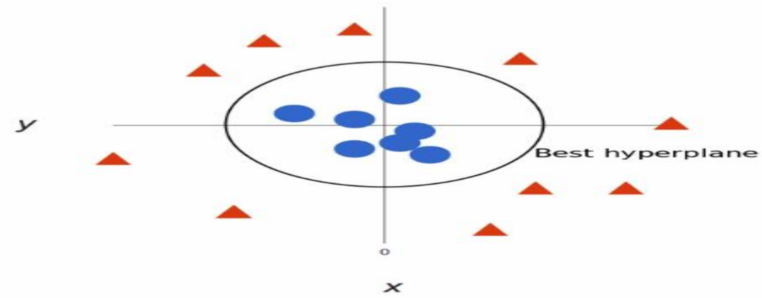
- The optimal hyperplane is the one with the largest distance between each tag. In two dimensions it looks like this :



- Those vectors are representations of your training texts, and a group is a tag you have tagged your texts with.
- As data gets more complex, it may not be possible to classify vectors/tags into only two categories. So, it looks like this:



- But that's the great thing about SVM algorithms – they're "multi-dimensional." So, the more complex the data, the more accurate the results will be. Imagine the above in three dimensions, with an added Z-axis, to create a circle.
- Mapped back to two dimensions the ideal hyperplane looks like this:



Metrics and Evaluation

- Cross-validation is a common method to evaluate the performance of a text classifier. It works by splitting the training dataset into random, equal-length example sets (e.g., 4 sets with 25% of the data). For each set, a text classifier is trained with the remaining samples (e.g., 75% of the samples). Next, the classifiers make predictions on their respective sets, and the results are compared against the human-annotated tags. This will determine when a prediction was right (true positives and true negatives) and when it made a mistake (false positives, false negatives).

- With these results, you can build performance metrics that are useful for a quick assessment on how well a classifier works:

- **Accuracy:** the percentage of texts that were categorized with the correct tag.

- **Precision:** the percentage of examples the classifier got right out of the total number of examples that it predicted for a given tag.

- **Recall:** the percentage of examples the classifier predicted for a given tag out of the total number of examples it should have predicted for that given tag.

F1 Score: the harmonic mean of precision and recall.

Application

In this section, we explain using a high-resolution smartphone camera to identify and classify bird information based on deep learning.

Software Implementation

The vision-based bird detection system is completely based on developing the software for bird detection and classification. Therefore, no hardware components have been used in this project. The developed software implement three methods, described further in this chapter.

The software code has been developed by using a Python programming language. We have used the Google Colab notebook that runs in the cloud to train the images.

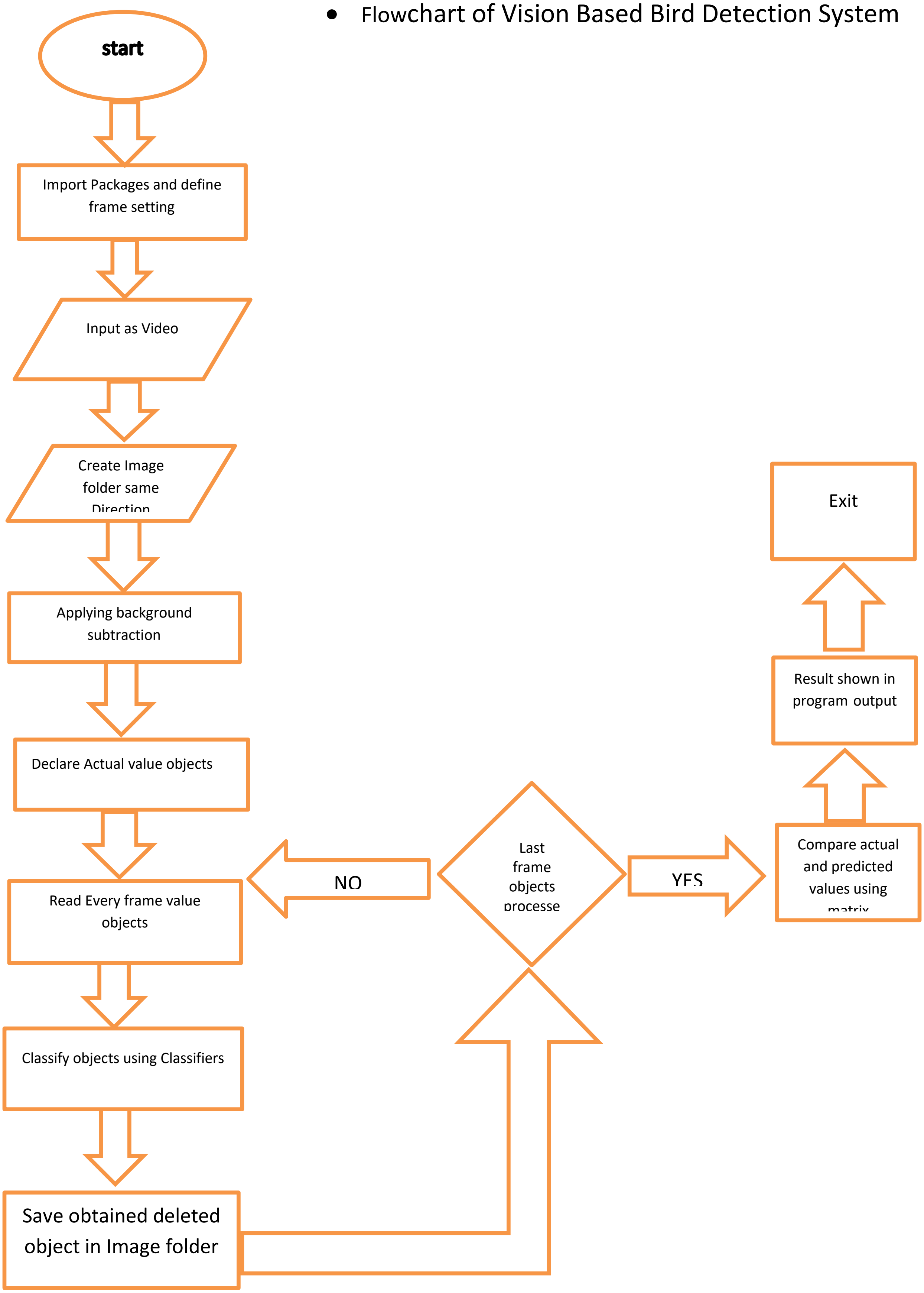
Flowchart and its Working

It involves the information from giving the input video to the programmed software to obtaining the desired result. It consists of the data briefly in sequential steps. The videos that are provided by the Bioseco company have been given as input.

The video will be converted into frames, where every frame will store in a folder with the same directory as the video. The frames will be then processed by background subtraction algorithm where the constant background frame is removed from the grayscale frames which, lets us determine the contour detection.

Contour detection is applied to the processed grayscale frames where a bounded box is formed around the desired moving objects. Now, these contour detected frames will be passed through, where the classification begins. The actual (manual) values have been declared in the program to let the software compare the actual and predicted values at a later step.

- Flowchart of Vision Based Bird Detection System

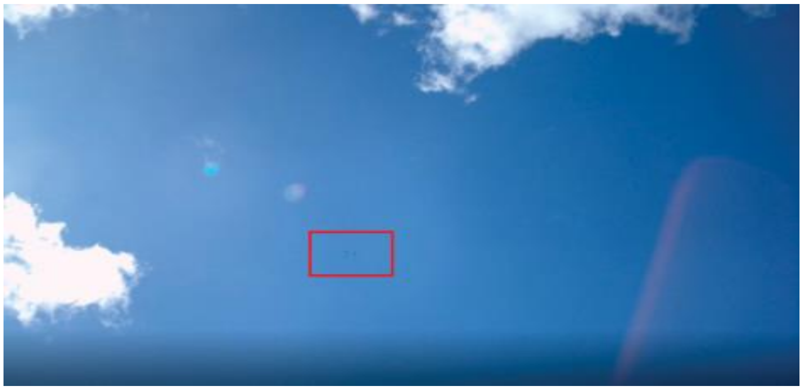


The classification begins with reading the objects in every frame. The detected objects were classified from the initial frame to the last frame. In the processed frames, the detected objects will be further marked with labels. These frames store the added labels in data that is created by the program. This data is called the predicted(y-pred) data. This obtained object data will be stored in the image folder.

This processing of frames will continue until the completion of the last frame. After obtaining the data from the last frame, comparing the actual data and the obtained data will begin by using the confusion matrix. The matrix compares the actual data values(y-true) and the obtained data values (y-pred) and create four data values such as TP, TN, FP and FN. The confusion matrix is obtained and the result can be seen in the program output.

Implementing Background Subtraction

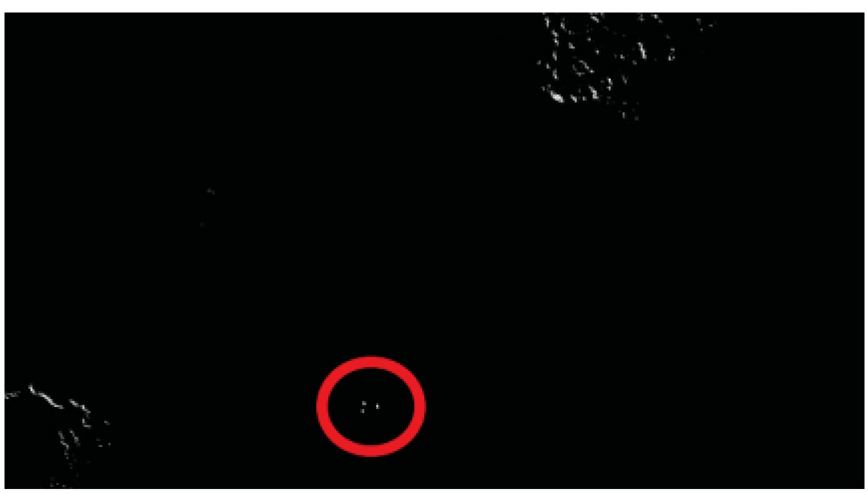
Background subtraction [47] is a method that is used for developing the foreground mask. We have used Gaussian mixture-based foreground/background subtraction in our software. We have first read the given input video or image sequences using " cv::VideoCapture ". After giving the input, we have created the background subtraction model by using "cv::BackgroundSubtractor class". In this subtractor, we have given three parameters to initialize the background subtraction, those are history, varThreshold, and detectShadows. We have given the history as "2" and [48] varThreshold as "10" and "False" for the detect Shadows to extract the high resolution and clear output of the video. Later, the output can be obtained by using the "imshow" function for the applied background subtraction.



Before Background Subtraction

Enlarged Image showing the birds

The extracted grayscale image from the given input video frame is shown. As one can see here, the three small dots together in the middle of the image are birds, which can be found after doing the background subtraction to the given input.



Implementing Contour Detection

After the background subtraction, we have used the contour detection method to create a bounding box around the object in the image with subtracted background. We have given this background subtracted output as an input to the contour detection and given the function "detections = []" to store those detections. After storing the detections, we will go through each detected object. Now we have given the "for" loop in contours. To calculate the area and remove the small elements, we have given [50] "cv2.contourArea(cnt)" which will count the area in pixels of an object. Now, we have given an "if" statement to detect the object with its area greater than 30 pixels and store the data in the detection list by giving"if (cont_ar > 30 and cont_ar < 20000):" [51]. By using this "cv2.boundingRect(cnt)" we have given four parameters to create a bounding box around the object. Here Figure 5.5 shows the output of the contour detection stage where the detected bird and blade are indicated by the bounding box.

Contour Detection

Conclusion

. In conclusion, implementing bird image retrieval and recognition using deep learning has proven to be a promising and effective approach. The utilization of convolutional neural networks (CNNs) allowed for robust feature extraction, enabling accurate identification of bird species. The retrieval system, coupled with image recognition, exhibited commendable performance in handling diverse datasets.

Furthermore, the model's ability to generalize across various bird species and adapt to different environmental conditions highlights its versatility. The incorporation of transfer learning, leveraging pre-trained models, notably improved efficiency and reduced the need for extensive labeled data.

Despite these advancements, challenges such as limited dataset diversity and potential biases in training data remain areas for improvement. Ongoing research and development efforts should focus on expanding the dataset and addressing potential biases to enhance the model's reliability in real-world scenarios.

In summary, the presented deep learning-based bird image retrieval and recognition system showcase promising outcomes, providing a foundation for future advancements in avian species identification and contributing to biodiversity research and conservation efforts.

REFERENCES

- https://archive.epa.gov/water/archive/web/html/birds-initiatives.html.
- https://www.gov.uk/government/publications/wild-birds-licence-to-killor-take-for-conservation-purposes-gl40/list-of-endangered-woodlandbirds.
- Mirugwe A, Nyirenda J, Dufourq E. Automating Bird Detection Based on Webcam Captured Images using Deep Learning. InProceedings of 43rd Conference of the South African Insti 2022 Jul 18 (Vol. 85, pp. 62- 76).
- Lynda Ben Boudaoud, Fr´ed´eric Maussang, Ren´e Garello, and Alexis Chevallier. Marine bird detection based on deep learning using highresolution aerial images. In OCEANS 2019-Marseille, pages 1–7. IEEE, 2019.
- https://www.kaggle.com/datasets/gpiosenka/100-bird-species.
- https://ebird.org/home
- https://www.sciencedirect.com/science/article/abs/pii/S0003682X19303482
- https://www.sciencedirect.com/science/article/pii/S2352340920301955
- https://developers.google.com/machine-learning/guides/text-classification
- https://monkeylearn.com/what-is-text-classification/
- https://www.allaboutbirds.org/news/whats-that-bird-song-merlin-bird-id-can-tell-you/
- https://ieeexplore.ieee.org/document/8524677
- https://dcase.community/challenge2018/task-bird-audio-detection
- https://www.birds.com/blog/identifying-birds-by-sound/
- https://birdnet.cornell.edu/
- https://github.com/kahst/BirdNET-Analyzer