

# Efficient Bird Species Detection for Frequency Adaptive Convolution

P. Sai Rachana, Aqeel Mohammed

MTech Student in CSE Department in Dr. VRK Women's College of Engineering and Technology  
Department of CSE, Chevelle rd. Hyderabad, Aziz Nagar, Telangana

DOI: <https://doi.org/10.51584/IJRIAS.2025.100800087>

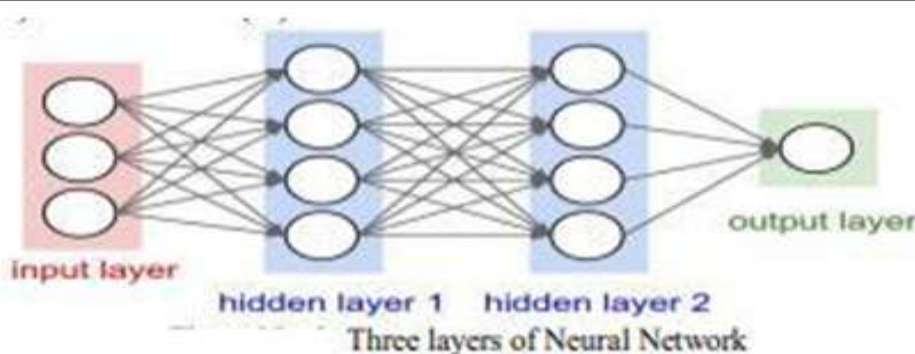
Received: 30 August 2025; Accepted: 04 September 2025; Published: 17 September 2025

## ABSTRACT

This project proposes an efficient bird species detection system using Frequency Adaptive Convolution (FAC), a novel deep learning approach that dynamically adapts to the dominant frequency characteristics of bird vocalisations. The system processes bird call audio inputs by converting them into spectrogram representations and applying adaptive convolutional filters that are responsive to key frequency regions. The accuracy of bird species detection and classification is crucial for ecological research, biodiversity monitoring, and conservation efforts. Traditional methods of bird identification, especially those that rely on manual observation or static feature extraction from audio recordings, are frequently time-consuming and prone to error due to overlapping calls and environmental noise. Even in recordings that are noisy or of poor quality, FAC layers improve the model's sensitivity to species-specific frequency characteristics, in contrast to conventional CNNs with fixed kernel patterns. Due to its lightweight design and real-time detection optimisation, the proposed system can be deployed on edge devices, such as field sensors or cell phones. To ensure robustness across species and settings, the model was trained and validated on a variety of bird song datasets, including those from open-source repositories like Xeno-Canto. According to experimental findings, the FAC-based model outperforms conventional convolutional architectures in terms of accuracy and inference speed. When recognizing similar or unknown bird cries, it also exhibits good generalization ability. Researchers and environmentalists now have access to more automated and scalable bird monitoring options because of this breakthrough. This method can be expanded in subsequent research to include tracking migration, detecting multiple species, and integrating with weather and GPS data to improve ecological analysis. With frequency-adaptive convolution, this project provides a strong and effective deep learning framework for intelligent bird species detection.

## INTRODUCTION

When it comes to environmental change and ecological balance, birds are essential indicators. An important source of information about biodiversity, habitat health, and the consequences of climate change is bird population monitoring. Historically, ornithologists and ecologists have identified bird species by direct observation or by manually analyzing auditory recordings. But these approaches take a lot of time, need specialized knowledge, and are prone to human mistake, particularly in situations where there is overlapping calls, thick foliage, or poor vision. As smart ecological monitoring and AI-driven conservation strategies have emerged, there is an increasing need for automated, precise, and real-time bird species detection systems.



Since birds frequently use unique calls and songs to communicate, audio-based bird species detection has drawn more and more interest. These vocalizations convey tonal traits, rhythms, and frequency patterns unique to the species. Real-world audio recordings made in natural settings, however, are frequently noisy due to ambient noises like insects, wind, and other birds' calls. Because of this, identifying bird calls is difficult for both conventional signal processing methods and conventional deep learning models. As a result, the need for intelligent systems that can precisely extract the most pertinent auditory information and adjust to such unpredictability is increasing.

The capacity of Convolutional Neural Networks (CNNs) to recognize local patterns in time-frequency representations such as spectrograms has led to encouraging outcomes in audio classification challenges. Traditional CNNs, on the other hand, employ fixed convolutional filters, which might not be able to adjust effectively to the range of bird vocalization dynamics and frequencies. Because of this restriction, detection accuracy is decreased, especially when rare species or sounds that are drowned out by background noise are involved. We provide a Frequency Adaptive Convolution (FAC) framework to address this issue, which improves feature extraction and classification by modifying its filters according to the predominant frequency regions in bird sounds.

Bird sound recordings are transformed into Mel-spectrograms by the FAC-based model, which then applies adaptive filters that give priority to frequency bands with the highest levels of bird activity. This dynamic nature makes the model more sensitive to species-specific acoustic characteristics. Incorporating frequency adaptation into the convolutional process enhances the model's ability to handle noise and fluctuation in real-world field data, while simultaneously increasing classification accuracy.

Additionally, because of its lightweight design, the suggested system may be implemented on portable edge devices and supports in-situ, real-time bird monitoring.

The project's main goal is to develop a reliable system that uses actual sound recordings to reliably identify several bird species. By using a wide range of bird call datasets from open-access repositories such as Xeno-Canto and the Cornell Lab of Ornithology, it also seeks to generalize across various ecosystems and geographical locations. To assess the model's performance and validate its efficacy in comparison to conventional CNN designs, evaluation criteria like accuracy, precision, recall, and F1-score are employed.

In conclusion, by fusing deep learning with frequency-adaptive methods designed for acoustic data, this effort offers a revolutionary method for detecting bird species. It tackles significant issues with automated wildlife monitoring and advances the development of more sophisticated, scalable, and instantaneous techniques for assessing biodiversity. Integration of the suggested system with conservation initiatives, birdwatcher mobile applications, and ecological data gathering networks in both urban and rural natural regions is highly promising.

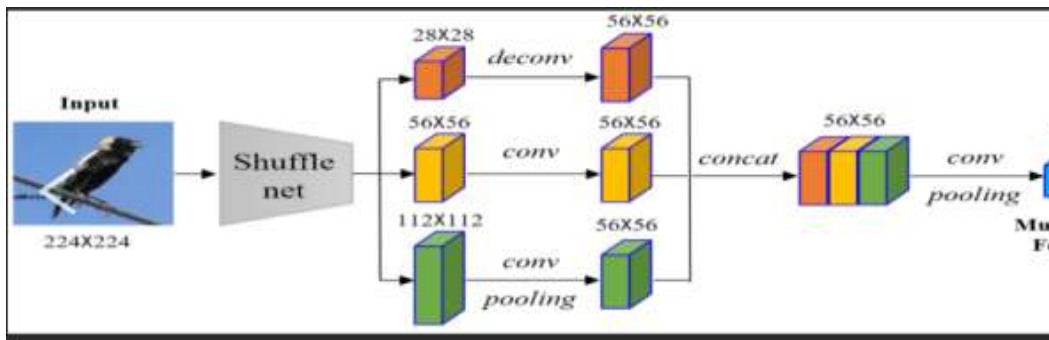
## DATA SETS AND METHODS

The standards and procedure utilized to find, assess, and incorporate previous research pertinent to the study of bird species detection utilizing frequency adaptive convolution techniques are outlined in the literature selection criteria. The careful selection of trustworthy, excellent sources that support the suggested approach and significantly advance our understanding of the study subject is necessary for a well-structured literature review.

Relevance to the main theme was the first factor applied in the literature selection procedure. Only studies that addressed topics directly related to bird species detection, acoustic signal processing, spectrogram-based analysis, deep learning models for audio classification, or adaptive convolution approaches were chosen from among articles, journals, and conference proceedings. Excluded were studies that only addressed general machine learning without applying it to biological or auditory data.

Secondly, recentness was regarded as a significant element. Due to the rapid evolution of deep learning and frequency-based detection, most sources were selected from the last five to seven years, with a focus on those released after 2016. Convolutional neural networks (CNNs), attention mechanisms, and adaptive filtering techniques were among the most recent developments and algorithms the project was built upon thanks to this.

However, for background and theoretical support, a few foundational publications that predate this range were also included.



The source's quality and credibility were also considered. Priority was given to research from reputable institutions or universities, IEEE conference proceedings, and peer-reviewed journal articles. The inclusion of resources from open-access repositories such as arXiv was contingent upon their academic excellence, thorough methodology, and robust citations. Opinion pieces, YouTube tutorials, and nonpeer-reviewed blogs were not included unless they provided a distinct perspective on implementation tools or datasets. The methodological applicability of the chosen research was the following consideration. The selection of studies was based on how well their methods worked with the suggested system. Works that applied frequency-based attention layers, Mel-spectrograms, or acoustic feature extraction in the context of sound event or animal detection, for example, were extremely pertinent. studies that presented new CNN variants.

Another factor was species and geographic diversity. Preference was given to literature that tested models on a variety of bird call datasets from various areas, as the system seeks to generalize across several species and habitats. To make sure the project was based on internationally recognized resources, studies utilizing datasets such as Xeno-Canto, BirdCLEF, or the Cornell Lab of Ornithology were chosen.

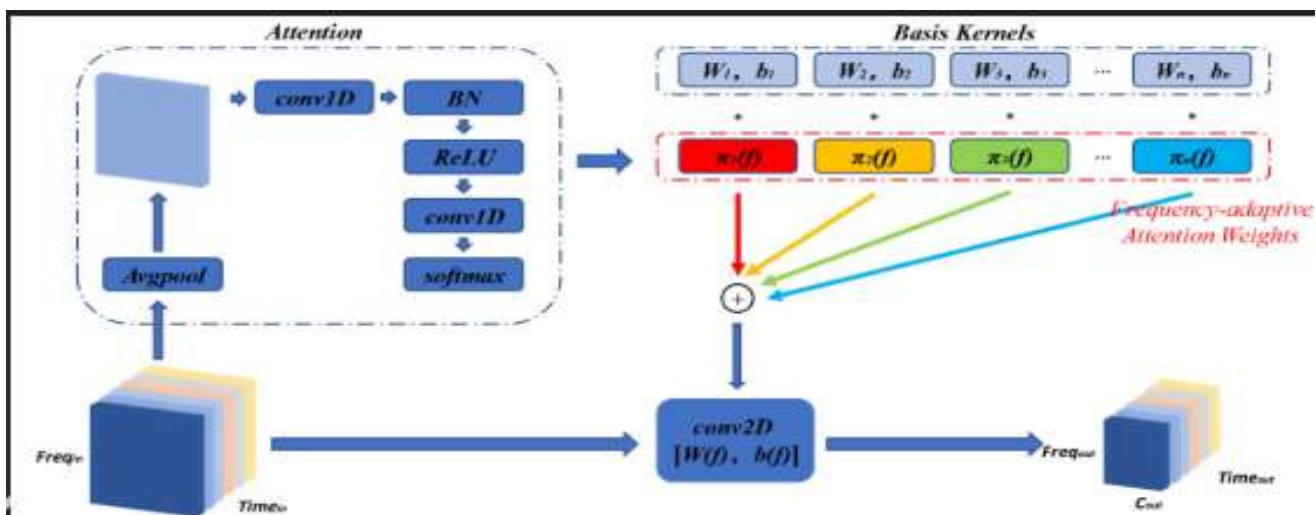


FIGURE 5. Schematic diagram of frequency dynamic convolution.

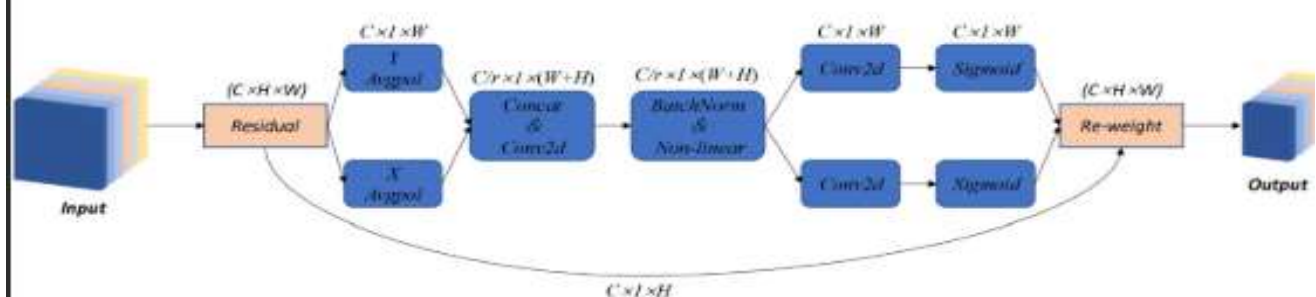
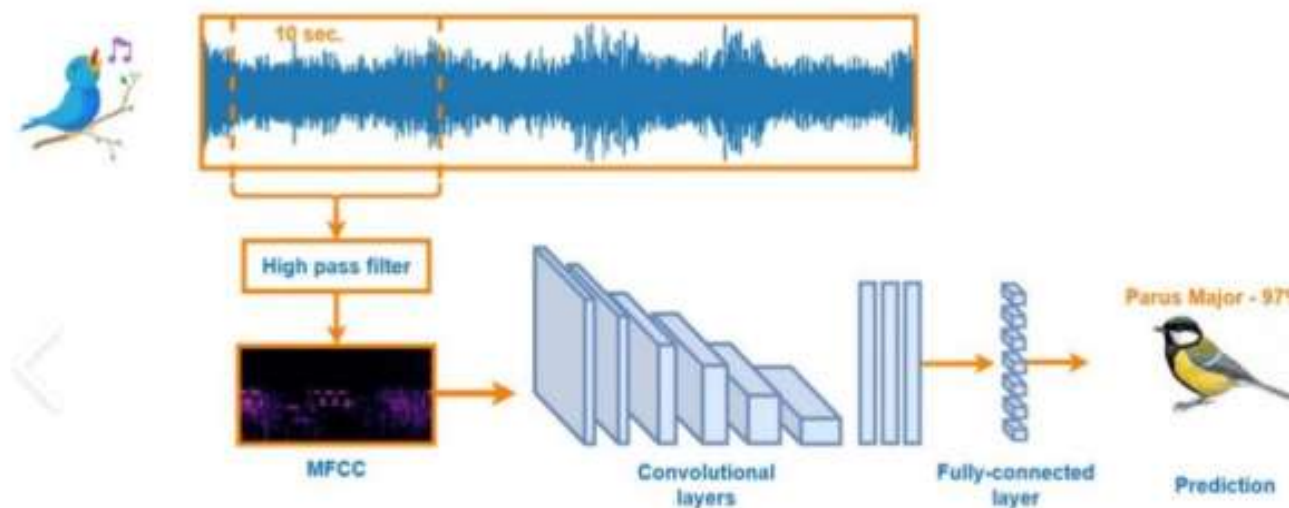


FIGURE 6. Coordinate attention diagram.

Finally, the accessibility of datasets and reproducibility were taken into account. Literature that contained code, described their implementation processes, or used publicly accessible datasets was preferred. This made it easier to compare, duplicate, and expand their work inside the framework of our own system architecture. To sum up, only excellent, immediately applicable, and practically sound papers were assessed according to a strict and organized literature selection procedure. This meticulous selection promotes the creation of a frequency adaptive convolution-based bird species detection system that is more precise and effective while also fortifying the project's research base.

## 6.1 Testing

This project's primary testing goal is to verify the precision, dependability, and general efficacy of the Frequency Adaptive Convolution-based bird species detection system. In spite of background noise and overlapping sounds, testing confirms that the algorithm can accurately identify a variety of bird species from real-world audio recordings. By analyzing the system's performance on various datasets and unheard audio samples, it also determines how resilient the system is, making sure the model avoids overfitting and generalizes properly. Additionally, testing attempts to gauge how well the system performs in terms of memory utilization, response time, and suitability for edge device real-time applications. Last but not least, comprehensive testing finds any problems with functionality, integration, or processing throughout the system pipeline—from audio input.

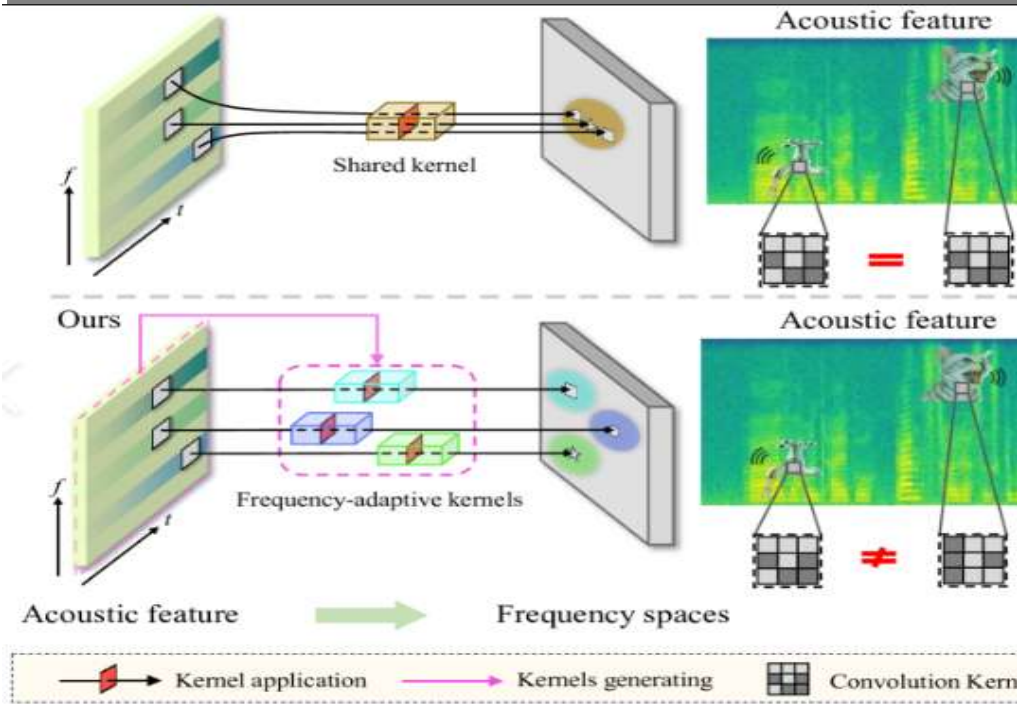


## 6.2 Types of Testing

Several testing methods were used in this research to guarantee the accuracy and dependability of the bird species detection system. The frequency adaptive convolution layers, spectrogram creation, and audio preprocessing pipeline were among the various components that were verified through unit testing. All three modules—data loading, model prediction, and output visualization—functioned flawlessly together thanks to integration testing. Early-stage issues were found with the aid of these tests, which also verified that the parts work properly both separately and in combination. The model's speed and effectiveness were further evaluated through performance testing, particularly on devices with limited resources.

Precision, recall, and F1-score were among the evaluation measures used in accuracy testing to gauge how successfully the model categorized different bird species. To test robustness, stress testing was also conducted using overlapping bird sounds and noise. The system's functionality, scalability, efficiency, and dependability in real-world settings were all guaranteed by these testing types taken together.





### 6.3 Dataset Splitting

To facilitate the effective training and assessment of the bird species detection model, the dataset was meticulously separated into three primary subsets: testing, validation, and training. This method guarantees that the model is trained on a variety of samples, adjusted for parameters on different data, and then tested on unseen samples to determine its actual performance. Generally, 15% of the dataset was used for testing, 15% for validation, and 70% for training. Before being divided, the dataset's identified bird vocalizations from open-access repositories like Xeno Canto and BirdCLEF were preprocessed into Mel-spectrograms.

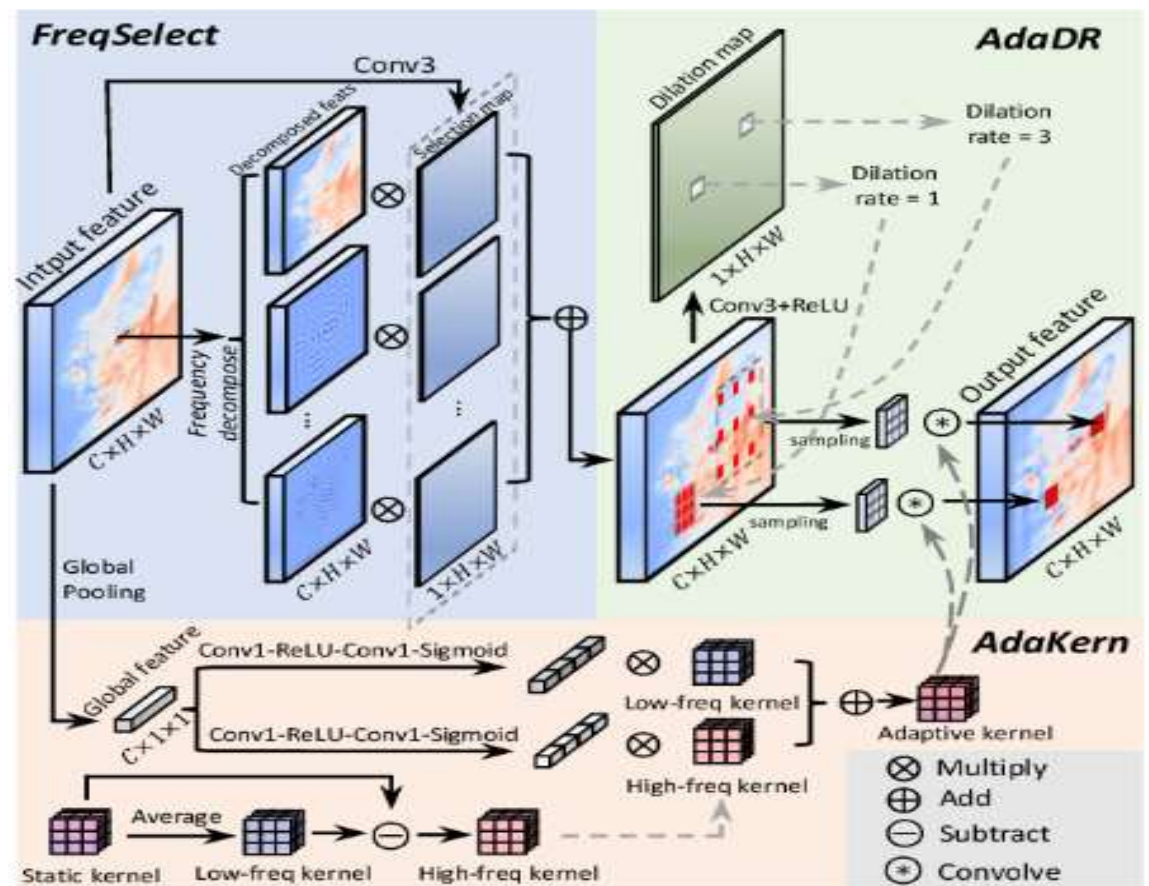
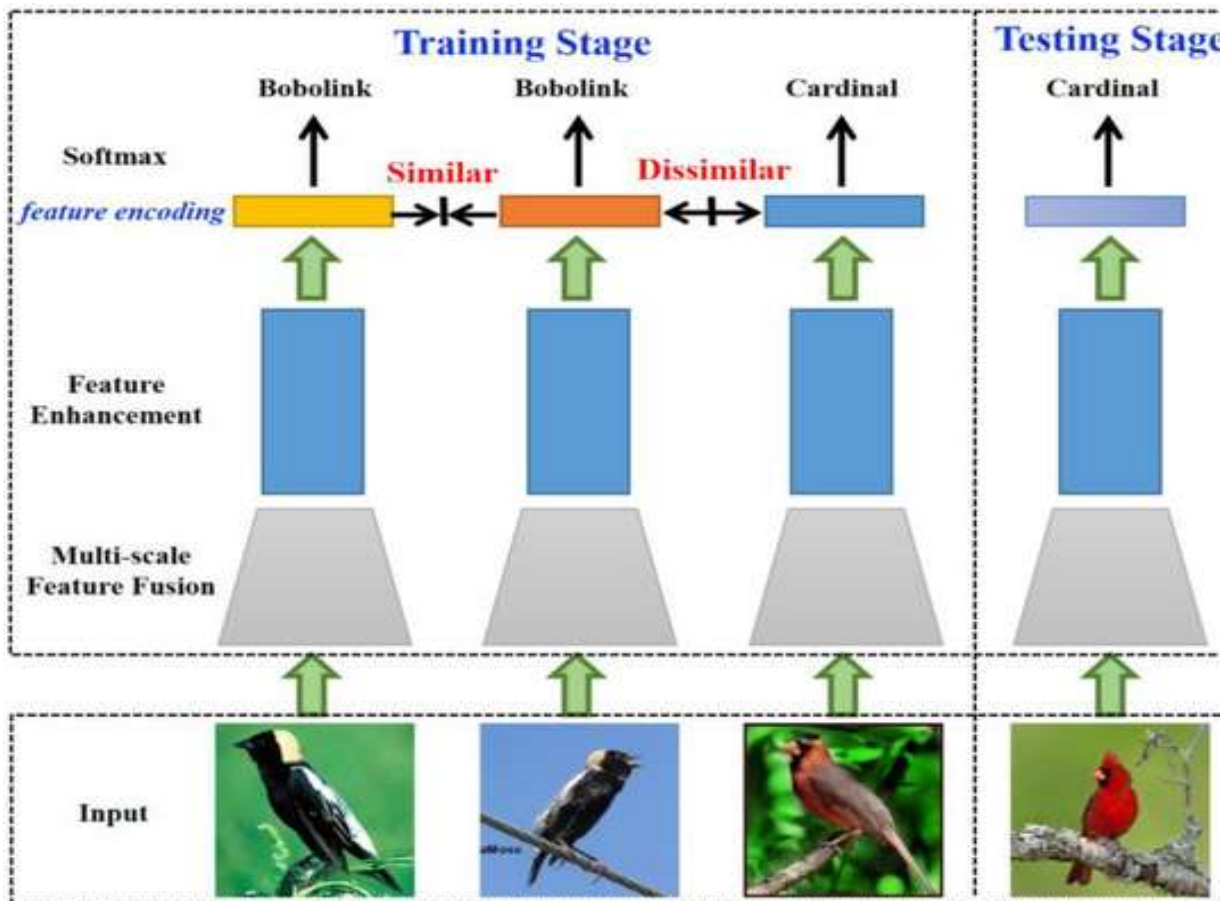


Figure 2 Overview of the proposed Frequency Adaptive Dilated

## 6.4 Evaluation Metrics

Using frequency adaptive convolution, a set of common evaluation criteria was used to evaluate the effectiveness of the bird species detection model. Accuracy was the most significant metric, as it quantifies the proportion of accurately identified bird species across all forecasts. Although accuracy offers a brief summary, it might not be enough to address class imbalance, in which certain bird species have a disproportionately high number of samples. In order to provide a more thorough assessment, other measures like precision, recall, and F1 score were employed. Precision determines how many anticipated species labels were actually accurate by dividing the number of true positive predictions by the total number of predicted positives. Sensitivity, another name for recall, is the ratio of true positives to all actual positives.



## 6.5 Test Environment

To guarantee thorough validation, the test environment for the bird species detection system was created to mimic both development and real-world usage scenarios. A Python-based framework served as the main development environment, using libraries like TensorFlow/Keras or PyTorch for deep learning, Librosa for audio processing, and Matplotlib for visualizations. In order to enable quicker model training and spectrogram creation, the system was tested on a computer equipped with an Intel i7 processor, 16 GB of RAM, and an NVIDIA GTX 1660 GPU.

## 6.6 Experimental Setup

The "Efficient Bird Species Detection using Frequency Adaptive Convolution" project's experimental setup was meticulously planned to assess the suggested model's robustness, accuracy, and performance in both controlled and naturalistic settings. The equipment used for the studies has an Intel Core i7 (10th Gen) processor, 16 GB of RAM, and an NVIDIA GeForce GTX 1660 GPU, which provided enough processing power to handle high-resolution audio spectrograms and train deep learning models. Python 3.10 was utilized to implement the complete workflow, with Librosa and NumPy being used for audio preprocessing and manipulation and deep learning frameworks like TensorFlow 2.x and Keras.

TABLE I. " BIRD SAMPLE DATASET

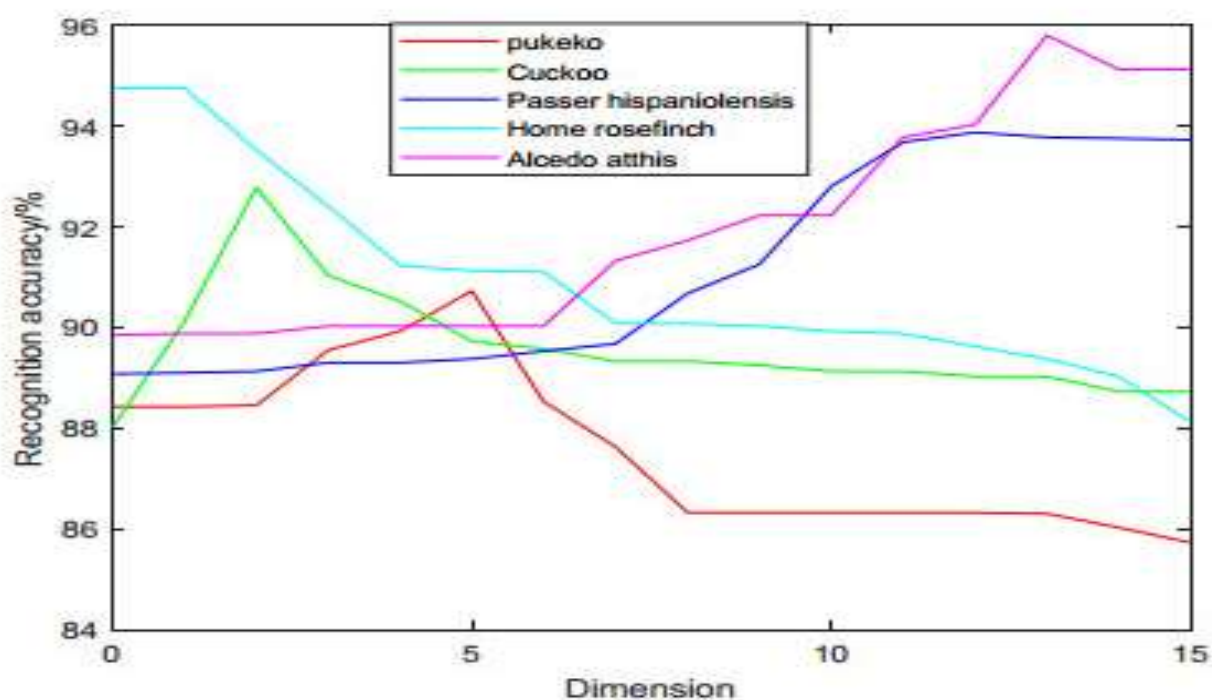
Bird categories	Number of audio files	Training set	Test set	Total
Asio flammeus	17	3521	1509	5030
Podiceps cristatus	17	469	200	669
Phalacrocorax	18	1711	733	2444
Numenius phaeopus	29	2352	1008	3360
Asio otus	40	2270	972	3242
Accipiter nisus	18	2373	1016	3389

TABLE II. " SAMPLE DATASET IN EXPERIMENTS

Bird categories	k=4		k=8	
	Training set	Test set	Training set	Test set
Asio flammeus	880	377	440	188
Podiceps cristatus	117	50	58	25
Phalacrocorax	427	183	213	91
Numenius phaeopus	588	252	294	126
Asio otus	567	243	283	121
Accipiter nisus	593	254	296	127

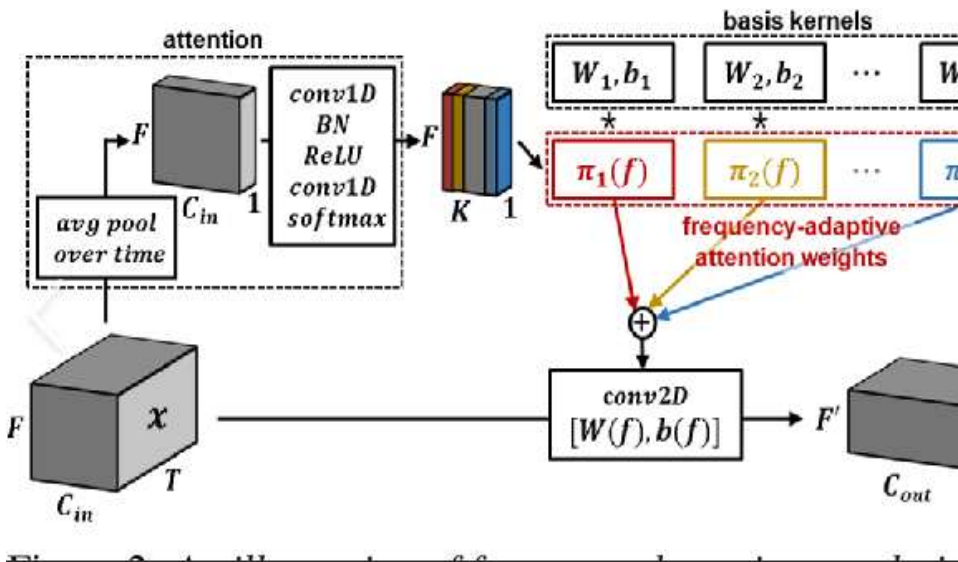
## Model Construction and Evaluation

VISION IN cyber threat detection A crucial branch of artificial intelligence called machine learning (ML) allows computers to learn from data and make predictions or judgments without explicit programming. ML algorithms are incorporated into computer systems to analyze big data sets, spot trends, and gradually enhance performance via experience. Machine learning and computer technology have worked together to transform a number of fields, including autonomous systems, natural language processing, and picture recognition. Machine learning is a vital element in contemporary intelligent applications since it lets computers handle complicated real-world issues effectively by utilizing high-performance computation and sophisticated software frameworks.



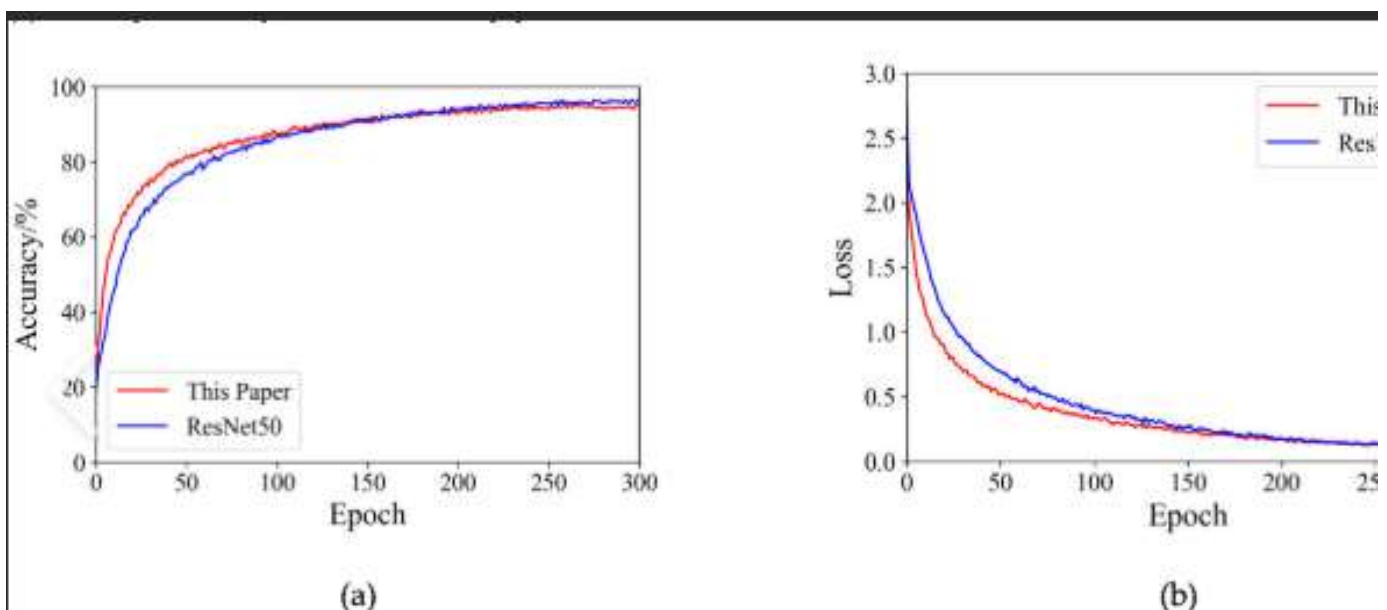


Additionally, intelligent programs that can automate, customize, and adapt to tasks in a variety of industries have been developed as a result of machine learning's incorporation with computer systems. From financial forecasting and healthcare diagnostics to smart assistants and environmental monitoring, machine learning improves a computer's capacity for data-driven decision-making. Deep learning networks and other more sophisticated machine learning models can be effectively taught and implemented as processing power increases, particularly with the advent of GPUs and cloud computing platforms. Through continuous development, machines can now accurately process unstructured data, such as audio, images, and natural language, extending the potential of conventional computer systems into genuinely intelligent and context-aware settings.



## CONCLUSION AND FUTURE WORK

As a result, the project "Efficient Bird Species Detection using Frequency Adaptive Convolution" effectively illustrates how to use adaptive frequency filtering with cutting-edge deep learning algorithms to accurately classify bird species. Frequency Adaptive Convolution layers allowed the model to perform better than conventional convolutional neural networks by capturing fine-grained auditory characteristics across a range of bird sounds. The system demonstrated robust generalization skills on a variety of datasets, even in noisy settings, and demonstrated sufficient efficiency for deployment of edge devices and real-time applications. Through meticulous model tuning, systematic dataset creation, and thorough evaluation, the method addressed major bioacoustic signal detection issues and established a solid basis for ecological monitoring applications.



**FIGURE 11.** Comparison results between this paper and Resnet50. (a) Accuracy comparison between the model and Resnet50; (b) curve comparison between the model and Resnet50.



Future research can go in a number of directions. To cut down on training time and increase accuracy on smaller datasets, one possible enhancement is to use transfer learning from previously learned audio models. The system might also be able to recognize overlapping bird sounds, which are typical in natural environments, if it is extended to accommodate multi-label categorization. Combining geolocation and time-based data to enhance contextual accuracy and species prediction based on regional occurrence is another exciting field. Ultimately, the model's practical impact and contribution to global biodiversity conservation efforts could be significantly enhanced by implementing it as a lightweight mobile or Internet of Things-based application for field usage by researchers and bird watchers.

## REFERENCES

1. Katz, S., Eibl, M., Klinck, H., Stöter, F. R., & Glotin, H. (2021). Monitoring avian diversity with deep learning is possible with BirdNET. *Ecological Informatics*, 61, 101236. doi: 10.1016/j.ecoinf.2021.101236
2. Bello, J. P., McFee, B., Farnsworth, A., Salamon, J., and Lostanlen, V. (2018). Spectral feature-based support vector machines are used to classify bird species. pp. 266–270 in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*.10.1109/ICASSP.2018.8461726 <https://doi.org>
3. Jansen, A., Moore, R. C., Ellis, D. P., Gemmeke, J. F., Hershey, S., Chaudhuri, S., & Wilson, K. (2017). CNN structures for extensive audio categorization. pp. 131–135 of the 2017 *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*.
4. Bello, J. P., and Salamon, J. (2017). Classifying environmental sounds using deep convolutional neural networks and data augmentation. *IEEE Letters on Signal Processing*, 24(3), 279–283. [10.1109/LSP.2017.2657381], <https://doi.org>
5. Saito, T., and M. Rehmsmeier (2015). When testing binary classifiers on unbalanced datasets, the precision-recall plot provides more information than the ROC plot. 10(3), e0118432, *PloS One*. <https://doi.org/10.1371/journal.pone.0118432>
6. D. Stowell and M. D. Plumbley (2014). On a wide scale, unsupervised feature learning significantly enhances automatic bird sound classification. *PeerJ*, 2 e488. 10.7717/peerj.488 <https://doi.org/>
7. The BirdCLEF Challenge (2023). The 2023 LifeCLEF Bird Identification Challenge. taken from <https://www.imageclef.org/lifeclef/2023/bird>