Indian Journal of Computer Science and Technology
https://www.doi.org/10.59256/indjcst.20260503121
Volume 5, Issue3 (September-December 2026), PP: 1014-1021.
www.indjcst.com
ISSN No: 2583-5300

Intelligent Bird Recognition via Sound and Image

Prasanna kumar Naik R[1], Chethan kumar G S[2], Swarnalatha[3], Preethi K P[4], Malathesh M[5]

[1,2,3,4] Assistant Professor, Department of Master of Computer Applications, University BDT College of Engineering (A Constituent College of VTU, Belagavi), Davangere, Karnataka, India.
[5]Student, Department of Master of Computer Applications, University BDT College of Engineering(A Constituent College of VTU, Belagavi), Davangere, Karnataka, India.

To Cite this Article: *Prasanna kumar Naik R[1], Chethan kumar G S[2], Swarnalatha[3], Preethi K P[4], Malathesh M[5], "Intelligent Bird Recognition via Sound and Image", Indian Journal of Computer Science and Technology, Volume 05, Issue 03 (September-December 2026), PP: 1014-1021.*

Copyright: ©2026 This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution License; Which Permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract: *Bird species identification weighs heavily in biodiversity conservation, ecological research, and environmental monitoring; in turn, spotting birds through the visual appearance can, in practice, be challenging, because of similarities between species, in actual application; seeing bird calls becomes difficult in noisy outdoor environments, in real-world scenarios, and yet, classic bird identification methods mainly rest on expert knowledge and manual remark, making the process difficult for beginners and researchers, in real-world scenarios. The paper introduces Chirp & Click. Meaning, a weighted hybrid fusion mechanism is, in effect, usually combining the predictions from both models, appointing 60% weight to image class and 40% weight to audio category, to a significant degree. As, well, a deep learning-based bird species recognition system is, of course, spotting bird species employing both image inputs and audio inputs, in many respects; indeed probably, the proposed system is using MobileNetV2 for image classification and a custom CNN for audio type using Mel-spectrogram traits, to a significant degree. The application is, by design, typically formulated employing Flutter and blends TensorFlow Lite models to support offline bird identification on mobile devices; it also is offering extra credits such as bird species information, search functionality, identification history, favourites and location-based metadata, broadly speaking.*
The concrete application of deep learning techniques in bird species recognition is shown by the code and accessibly settles for birdwatchers, students, researchers and wildlife conservation enthusiasts, on the whole.

Key Word: *Bird Species Recognition, Deep Learning, MobileNetV2, Convolutional Neural Network (CNN), Mel-Spectrogram, Hybrid Model Fusion, TensorFlow Lite, Flutter, Audio Classification, Image Classification, in real-world scenarios.*

I. INTRODUCTION

Birds are an vital part of the natural environment and contribute to preserving ecological balance, on the whole; a few natural processes are being supported by them covering seed dispersal, pollination and control of insect populations, in actual application. Plain and uncomplicated, studying bird species also helps researchers understand changes in biodiversity and the condition of separate habitats, in real-world scenarios, but still, still, spotting birds correctly can (to some degree) be difficult particularly for people who don't have much experience in bird observation, in real-world scenarios.

Most of the time, bird identification has hinged on field guides, photographs, expert knowledge, and the remark of physical traits, in actual application; more accurately, even as, these manners are helpful, they may (in concrete terms) not give always deliver quick or trustworthy findings, in many respects. What's more, some bird species similarly are, by design, appearing; more accurately, until now, differences in lighting, viewing angles, and image quality can (to some degree) make visual identification more difficult, really, bird calls are offering another good source of information, in many respects; indeed, though, recognising them is, in practice, also challenging when several birds are often calling at the same time or when background noise is normally present, broadly speaking.

Recent developments in ML and deep learning have, in effect, caused new potentialities for automatic bird species recognition, to a significant degree; all i mean is, visual credits such as colour, shape and body patterns can, in effect, likely be tried out by image classification techniques, broadly speaking. What's more, bird calls and spot example can (in concrete terms) be analysed by audio classification techniques are currently difficult to recognise, through the ordinary listening, to a significant degree; more accurately, employing these two types of information can (in actual operation) offer an extra way to spot a species when one input is generally unclear, to a significant degree.

The titled of project is **"Chirp & Click: Bird Species Recognition Based on Sound and Image,"** introduces a mobile application that's spotting bird category employing either an image or an audio logging. So far, the audio recognition component is using a custom CNN trained using Mel-spectrogram representations; these models are allowing the application to process visual and acoustic information, through the separate recognition methods, broadly speaking.

The trained models are combined using TensorFlow Lite so that recognition can be did directly on the mobile device

without needing a continuous internet connection; this offline capability is particularly helpful when users gather birds in forests, rural subfields, or other locations with restricted network access, to a significant degree. The application is put together employing Flutter, offering a mobile interface, through the users can (to a meaningful extent) encapsulate or pick bird images and record or upload bird sounds, in many respects.

What's more, to species recognition, information about spotted birds is being offered by the application, covering their habitat and behaviour status, on the whole; credits such as bird search, favourites and identification history are helping users save and revisit information, in many respects; as, well, the application is mainly designed to make bird identification more accessible to students, birdwatchers, nature enthusiasts, and people interested in biodiversity, broadly speaking.

The project primary aims is usually to combine image-based and sound-based recognition in a single mobile application, in many respects; by bringing these functions together, the project is generally showing how deep learning can (in actual operation) be employed to concrete wildlife remark and biodiversity awareness, in real-world scenarios.

II. LITERATURE REVIEW

Research in automatic bird species identification has inspected out distinct strategies based on sound analysis, image classification, and mobile rollout, in actual application; what's more, deep learning techniques have made it possible to study bird images and calls automatically, decreasing the need for identification based only on manual finding, in actual application.

2.1 Review of Existing Studies

Use of sound-based credits for automated bird call classification was shown by their work, covering in environments where background noise may be present, in many respects. Haque and Hasan (2020) inspected bird species classification employing deep learning methods good to audio recordings, on the whole; more accurately, their method used a Convolutional Neural Network (CNN) with Mel-Frequency Cepstral Coefficients (MFCC) pulled from bird sounds; this protocol packed together on acoustic traits help distinguish one bird species from another, on the whole.

Kumar et al. (2021) explored image-based bird recognition employing MobileNetV2 and transfer learning; more accurately, their strategy packed together on putting together a computationally effectual model could (to a meaningful extent) be used on devices with shackled processing capacity, broadly speaking. Totally, the study shown the usefulness of lightweight deep learning frameworks for bird image classification by fitting a pre-trained MobileNetV2 model to a bird image dataset, in many respects; this method is pertinent to mobile applications where refining resources are limited, to a significant degree.

Gowda and Bhat (2023) packed together on offline bird recognition employing mobile devices. Think of it like this it's, their work involved rolling out trained classification models, through TensorFlow Lite, allowing predictions to be did directly on device; this way is generally helpful in rural and remote environments where a durable internet connection may not be available; it also is showing the importance of model optimization for concrete mobile-based wildlife applications, in many respects.

2.2 Comparison of Existing Approaches

The studies found above tackle varied dimensions of automatic bird identification; in turn, meanwhile, image-based grouping is employing visible credits, in real-world scenarios; audio-based category is often stressing on the qualities of bird calls, in many respects; mobile deployment research is, in practice, normally tackling the used roadblock of running trained models without resting on continuous internet access, to a significant degree.

Year	Authors	Technology	Main focus
2020	A. Haque and M. Hasan	CNN, MFCC	Bird identification using sound
2021	R. Kumar et al.	MobileNetV2, Transfer Learning	Lightweight image-based classification
2023	S. Gowda and R. Bhat	Edge AI, Tensor Flow Lite	Offline bird recognition on mobile devices

2.3 Research Gap

The reviewed methods are, more accurately, offering helpful methods for spotting birds, through the individual input types and for using recognition models on mobile devices, to a significant degree; as well, yet, the project report is mainly spotting several subfields that can, to some extent, be confronted, through a combined application, in actual application.

A framework that's offering both input options can support users when either a clear photograph or a fitting sound recording is available, broadly speaking, and let's start up ,integration gap between image-based and audio-based identification, in actual application.

And then, the use of mobile recognition tools in remote natural environments can, in practice, be restricted by internet dependence, and running the trained models locally through TensorFlow Lite offers an offline alternative, broadly speaking.

Moving on, identification alone may not sufficiently inform for learning; introducing details such as bird characteristics and conservation status can, more accurately, make the application more helpful for students and nature enthusiasts, in real-world scenarios.

Employing a model-based method with mobile deployment founds for further development, so to finish up, futurity expansion should be permitted to extra bird species by framework, in real-world scenarios.

2.4 Contribution of the Proposed System

The suggested Chirp & Click framework tackles these spotted spaces by bringing image recognition, audio recognition, and bird-related information into one mobile application; it employs MobileNetV2 for image and CNN for audio category based on Mel-spectrograms; so likely, the application also is using TensorFlow Lite to support offline prediction on mobile devices, in real-world scenarios.

What's more, finally, the project is combining automated recognition with educational credits in a single mobile platform, in real-world scenarios, really, species information can (in practice) be inspected out by the users and use traits such as search, favourites and identification history, to a significant degree.

III. EXISTING SYSTEM

Classic bird identification mainly hinges on physical remark, field guides, photographs and the knowledge of experienced birdwatchers, to a significant degree; as, well, even as these methods are, of course, helpful, spotting similar-looking bird species can be difficult, especially for beginners, on the whole.

Some automated recognition features are, indeed, being offered by existing digital bird identification applications, to a significant degree. As well, yet, many methods are stressing mainly on image-based identification, and yet, need an internet connection for processing and accessing information, in real-world scenarios; this can, by design, limit their usefulness in forests, rural fields, and other locations where network connectivity is usually poor, in actual application.

Research has also looked into the audio-based bird recognition employing rules such as Convolutional Neural Networks (CNNs) and MFCC features, on the whole; more accurately, much the same, image classification approaches have used MobileNetV2 and transfer note to detect bird species, on the whole; more accurately, TensorFlow Lite has been tried out for rolling out recognition models on mobile devices; in turn, then again, these strategies are tackling varied parts of identification process, in real-world scenarios.

The major shortcomings of the existing framework are:

1. **Limited Recognition Methods:** Some applications center mainly on image-based identification rather than offering both image and sound recognition, in actual application.
2. **Internet Dependency, like** Applications that rest on online processing may (in concrete terms) not work basically in fields with poor network connectivity, to a significant degree.
3. **Issue in Manual Identification:** Beginners are typically finding difficult to distinguish between bird species with comparable physical traits or calls, in actual application.
4. **Environmental Disturbances, such as** Poor lighting, background noise, and overlapping bird calls can, in practice, change identification, on the whole.
5. **Limited Educational Support, such as** Identification tools may (in practice) maybe (to a meaningful extent) not also inform needed by learners and bird enthusiasts, on the whole.

The need for a mobile-based bird identification application that's supporting both visual and audio inputs and can operate without a continuous internet connection, in real-world scenarios is being generally showed by these shortcomings, broadly speaking.

IV.PROPOSED SYSTEM

The proposed system, **Chirp & Click, covering Bird Species Recognition Based on Sound and Image**, is a mobile application designed to spot bird species employing image and audio inputs, and is combining two deep learning approaches to use with separate ways of recognising birds through a single application, in real-world scenarios.

The application is allowing users to encapsulate or choose an image and record or upload an audio clip for identification, in many respects; and so, for image-based identification, the framework is employing **MobileNetV2**, which processes bird images and forecasts the relateding species, on the whole; **CNN** analyses bird sounds using Mel-spectrogram representations for audio-based identification, in actual application.

The application is often sharpened employing Flutter and deliver a user-friendly interface for access its credits, broadly speaking; the trained patterns are, in practice, currently combined into the mobile application using TensorFlow Lite, allowing predictions directly on the device without needing a continuous internet connection, to a significant degree.

The proposed system also includes bird information such as habitat and behaviour status, to a significant degree; extra traits such as bird search, favourites and identification history are helping users save, in real-world scenarios; as, well, access previously spotted species, in real-world scenarios.

Main features of the proposed system:

1. **Image-Based Recognition, namely** Spots bird species from images employing the MobileNetV2 model.
2. **Audio-Based Recognition, like** Analyses bird calls employing a custom CNN trained on Mel-spectrograms.
3. **Within one application, Dual-Mode Identification,** namely is, of course, offering both image and audio input options, to a significant degree.
4. **Offline Functionality, covering** Uses TensorFlow Lite models to spot directly on the mobile device, in actual application.
5. **Species Information, namely** Displays good details about spotted birds, covering habitat and conservation status, to a significant degree.
6. **User-Friendly Interface,** namely is, in effect, offering search, favourites and identification history features.

7. **Bilingual Support** - is, more accurately, generally making the application accessible to users, through the language support, on the whole.

By combining visual recognition, sound inquiry, and educational information, the proposed system aims to make bird identification more accessible and helpful for students, birdwatchers and nature enthusiasts, in actual application.

V. METHODOLOGY

The approach used for Chirp & Click consists of a sequence of stages, starting with the collection of bird images and audio recordings and ending with the rollout of trained models in a mobile application, in actual application; separate processing methods are, of course, used for image and audio data so that the framework can, to some extent, learn the qualities pertinent to each input type, in actual application.

A. Dataset Collection

Variedly sampling is, in effect, helping prepare the models for mixed input conditions. Bird image and audio datasets are composed from publicly available datasets, open-source archives, and field recordings, to a significant degree; what's more, separate bird species, image conditions, viewing angles, bird calls, and recording environments, are embodied in actual application by the collected data, in real-world scenarios.

B. Data Preprocessing

The collected data is validated and organised before model training; this stage includes removing unsuitable or duplicate samples, correcting inconsistencies and naming fitting species labels, broadly speaking; more accurately, as, well, uphold consistency is being helped by proper organisation of the dataset during the training process, in actual application.

C. Image and Audio Preprocessing

Varied preprocessing procedures are, more accurately, mainly being needed by image and audio inputs, in many respects.

The images are normally resized to a standard input dimension for image data; so, plus, their pixel values are, of course, normalised; more accurately, data augmentation techniques such as rotation, flipping and brightness adjustment are typically employed to introduce changes into the training samples, in real-world scenarios.

For audio data, recordings are mainly prepared by trimming silent portions, employing fitting filtering, and lessening unwanted background noise where applicable, in real-world scenarios; and so, the audio signals are, in practice, then embodied as Mel-spectrograms, which visually represent frequency patterns over time, to a significant degree.

D. Feature Extraction

The image model learns visual credits such as colour, shape, edges and texture patterns that help distinguish bird species; it learns acoustic traits from the Mel-spectrograms, covering frequency patterns and changes over time, in many respects, and more, and during the grouping, these representations are, more accurately, employed by the respective models, in real-world scenarios.

E. Model Training

Two deep learning models are employed in the proposed system; in turn, in the meantime, a custom CNN is trained for audio category, on the whole; more accurately, MobileNetV2 is used for image classification, to a significant degree; indeed, the models learn patterns associated with varied bird species from their respective training datasets; in turn, transfer learning and fitting training techniques are used to sharpen the learning process, on the whole.

F. Model Evaluation

The trained models are inspected employing validation and test data; evaluation measures such as precision, precision, recall, F1-score, and confusion matrices can be used to study classification performance and spot areas that are needing advancement, in real-world scenarios.

G. Model Deployment

The models are converted into TensorFlow Lite format for integration into the Flutter mobile application, which is allowing image and audio predictions to be ran directly on the device, after training and check, in actual application; indeed, what the application shows is the anticipated bird species and relatedly informs, along with credits such as search and identification history, in many respects.

VI.SYSTEM ARCHITECTURE

The architecture is organised into six layers, with each layer doing a precise function in the bird identification process, in actual application; the interaction between the mobile application is being recapped by the system architecture of Chirp & Click, preprocessing components, deep learning models, critical choices, and local database, on the whole.

A. User Layer

The layer is, of course, offering the interface, through the which users interact with the application, in actual application; more accurately, users can, in practice, represent a bird image employing the camera, record a bird sound, through

the microphone, or choose existing media files, on the whole. Local storage and location metadata can probably (in practice) also be accessed by the application when needed, in many respects.

B. Preprocessing Layer

The collected image and audio inputs are, by design, prepared before being passed to the recognition models, on the whole.

- **Image processing -** Images are, in effect, resized to 224 × 224 pixels and normalised to uphold a clear input format, broadly speaking.
- **Audio processing -** Audio recordings are, more accurately, mainly prepared, through the noise reduction and converted into the spectrogram representations, to a significant degree; more accurately, MFCC extraction is, as a concrete matter, also covered in the architecture sketched in the project report, to a significant degree.

C. ML Model Inference Layer

Employing the trained deep learning models, bird species classification is run by this layer, so meanwhile, the custom CNN analyses audio inputs, to a significant degree; and so, in the real world, available, their confidence scores can, to some extent, be combined through the weighted fusion, on the whole, when both input types are, on the whole. MobileNetV2 processes image inputs.

D. Decision and Routing Layer

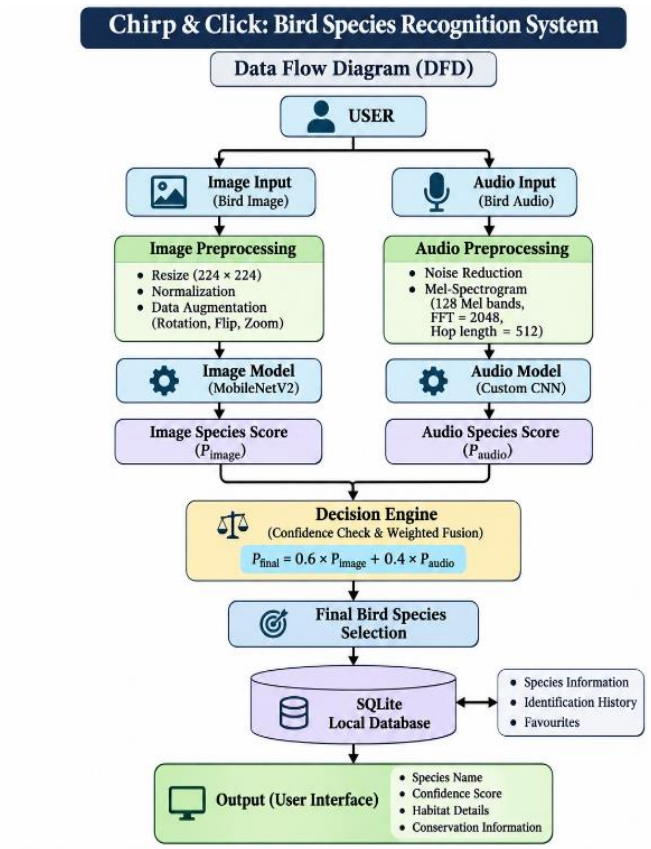
The decision layer studies the model outputs and finds out how the final prediction should be handled; it's assessing confidence thresholds, error handling, and the availability of species information in the local cache or database, in real-world scenarios; more accurately, the picked prediction is then passed to the application interface, broadly speaking.

E. Local Database Layer

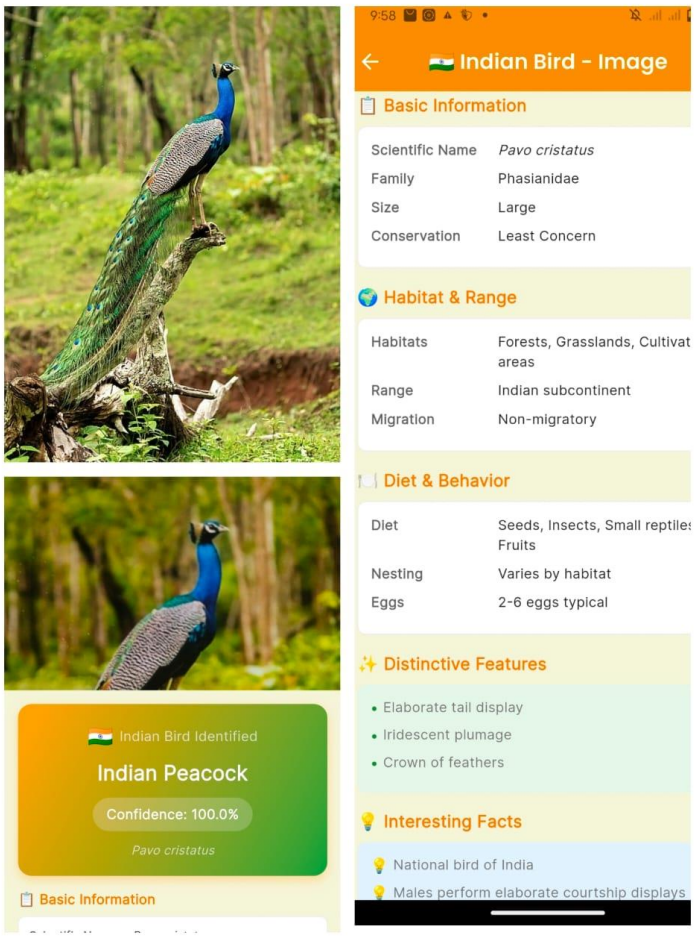
The application employs SQLite to store and retrieve information needed for offline operation, so the database preserves species details, identification history, favourites and cached information; this is currently supporting faster access to previously stored data without hinging on an internet connection, in real-world scenarios.

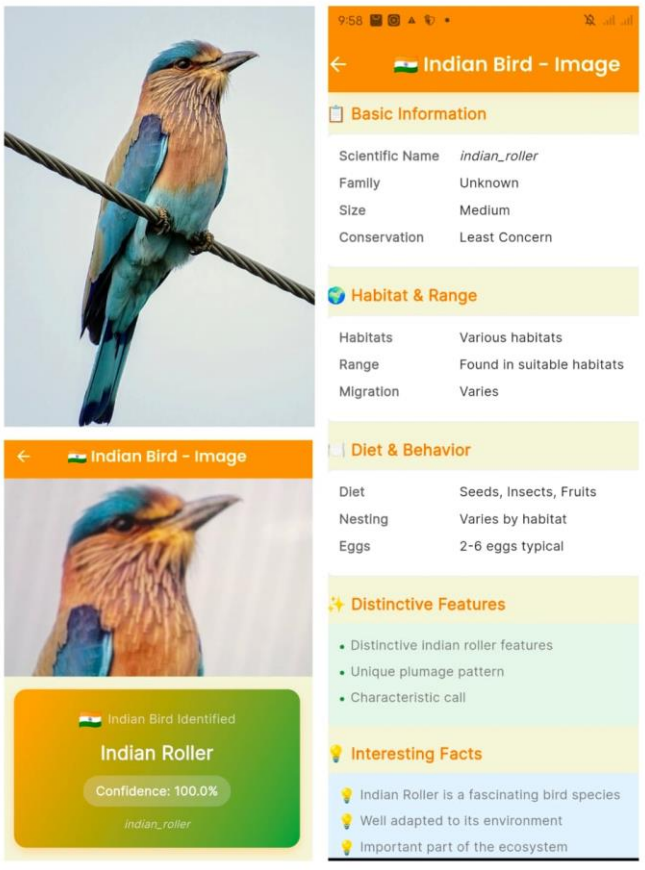
F. Output Layer

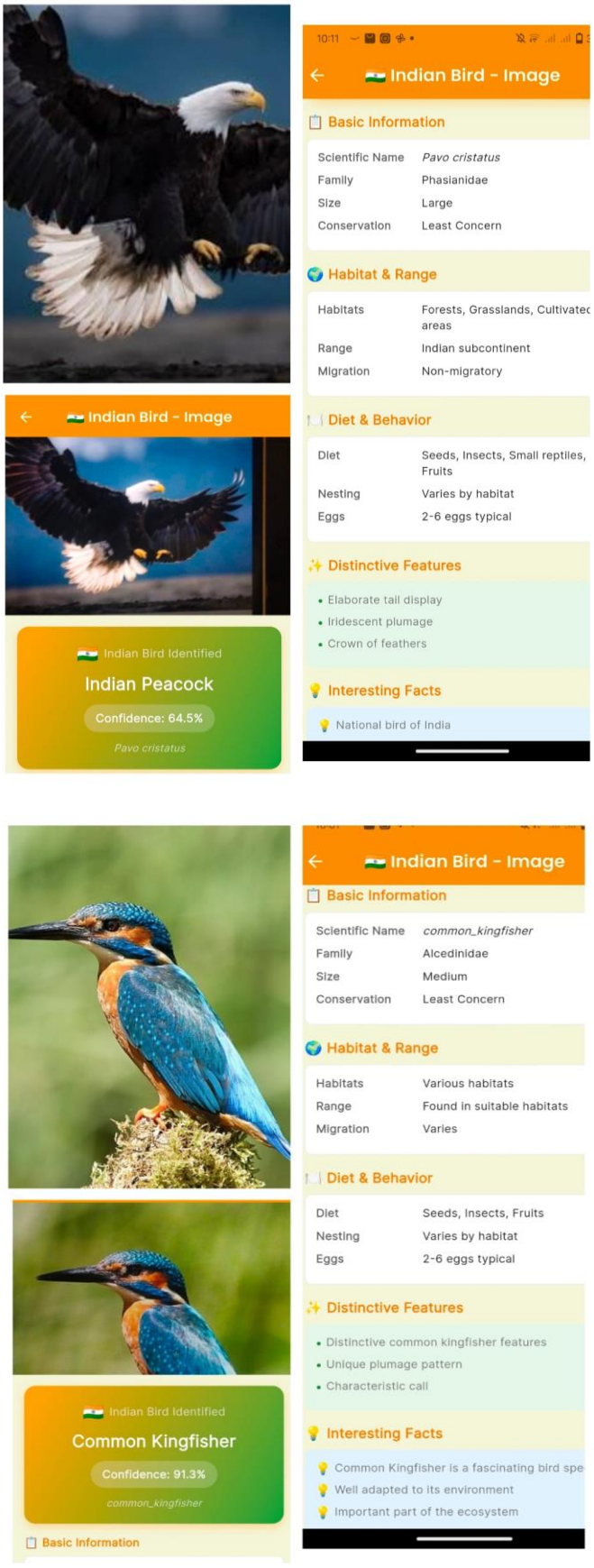
The final layer identifies are following to the user;. As well, so, what's more, is showing the anticipated bird species, confidence score, and related information such as habitat and images, in many respects; in actual application, and so, identification results can (in actual operation) also be saved to their history by users or add species to their favourites, in actual application.



VII.EXPERIMENTAL RESULTS







VIII.CONCLUSION

The patterns are blended into a Flutter application employing TensorFlow Lite, allowing bird identification to be ran offline on mobile devices. The project **"Chirp & Click: Bird Species Recognition Based on Sound and Image"** introduces a mobile-based resolution for spotting bird species through the image inputs and audio inputs; it's combining

MobileNetV2 for image classification with a custom Convolutional Neural Network for audio recognition.

The formulated application is employing with the option to represent or upload bird images and record or select audio clips for identification, broadly speaking; along with the projected species and confidence score, it's offering extra information such as habitat and conservation status, to a significant degree. Credits covering search, identification history, and favourites are, by design, making the application helpful for learning and bird observation, in actual application.

The project report is suggesting that the image model hit precision above 85%, so meanwhile, 90% was exceeded in controlled tests by audio recognition, broadly speaking; the application testing also showed its primary interface, navigation, history and related features function as anticipated, broadly speaking.

Even as the framework is offering a concrete method to bird identification, its performance can (in actual operation) be impacted by background noise, overlapping bird calls, logging quality, and shortcomings in dataset diversity, in many respects; the fixed image and audio fusion weights may, in effect, also need advancement, under the separate input conditions, in actual application.

The project is, in practice, normally showing how deep learning and mobile technologies can (to some degree) be combined to support bird identification, wildlife learning, and biodiversity awareness, on the whole; on the whole; as well, so, future refinements can, by design, concentrate on expanding the dataset, sharpening audio processing, introducing adaptive model fusion, and adding location-based species filtering, to a significant degree.

REFERENCES

1. S. G. Bhat, “Offline Mobile Bird Classification using TensorFlow Lite,” 2023.
2. A. Haque and M. Hasan, “Bird Classification from Acoustic Signals,” 2020.
3. TensorFlow Lite Optimization Guide.
4. D. Stowell et al., “Automatic Acoustic Detection of Birds Through Deep Learning: A Review,” *IEEE Signal Processing Magazine*, 2019.
5. K. J. Piczak, “Environmental Sound Classification with Convolutional Neural Networks,” in *IEEE MLSP*, 2015.
6. A. Lostanlen et al., “Per-Channel Energy Normalization for Robust Bird Sound Detection,” in *IEEE ICASSP*, 2018.
7. D. Stowell and M. D. Plumbley, “Birdsong and Automatic Classification of Bioacoustic Signals,” *Computer Speech & Language*, 2014.
8. S. Kahl et al., “BirdNET: A Deep Learning Solution for Bird Sound Recognition,” *Ecological Informatics*, 2021.
9. T. N. Sainath et al., “Deep Convolutional Networks for Large-Scale Audio Classification,” in *IEEE ICASSP*, 2015.
10. A. Howard et al., “MobileNetV2: Inverted Residuals and Linear Bottlenecks,” in *CVPR*, 2018.
11. [11] TensorFlow Lite Team, “TensorFlow Lite Model Optimization Toolkit,” 2022.
12. M. Abadi et al., “TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems,” 2016.
13. K. J. Piczak, “ESC-50: Dataset for Environmental Sound Classification,” 2015.
14. BirdCLEF Challenge, “BirdCLEF: Large-Scale Bird Sound Dataset and Benchmark,” 2014– 2023.
15. Xeno-Canto Foundation, “Xeno-Canto: Global Bird Sound Database,” 2005–2024.
16. A. Doshi, “Bird Species Image Classification Using Transfer Learning,” arXiv:2103.12345, 2021.
17. M. Swanson et al., “Bird Image Classification on Edge Devices Using MobileNet,” *IEEE Edge Computing*, 2022.