CS 719 Data Science Project

AvianNET

Intelligent Bird Species Identification System through Audio Recordings using CRNN

SUPERVISED BY
DR. HOWARD HAMILTON

PRESENTED BY
SUNNY CHOWDHARY- 200483456

Ideas Comparison: Audio vs. Visual

Which method of bird species identification works better, hearing or sight?

- Bird enthusiasts use visual and audio identification to recognize species.
- Audio identification excels at recognizing species by calls and songs.
- This project investigates audio identification's efficacy in species recognition.

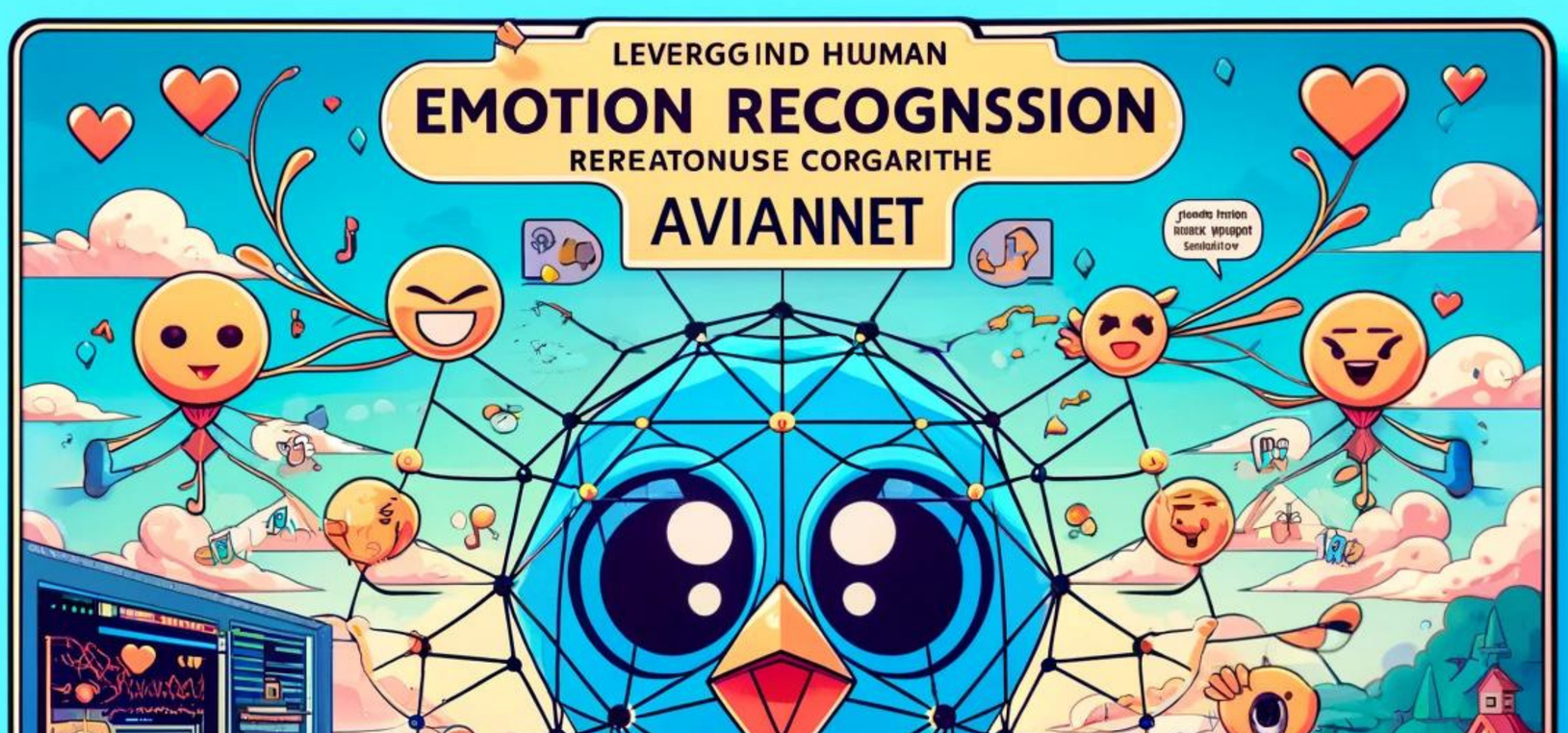
Identifying Birds By Sound

Learn the skill to identify birdcalls and songs.

- Project uses bird sounds to help birdwatchers identify various species.
- Users upload bird noises for algorithm-based precise identifications.
- Focus on identifying common North American birds.

Spark for Identifying Birds from their calls

Drawing from human emotion recognition research, our project uses CNNs and MFCC techniques for bird sound identification, emphasizing feature extraction and model optimization to accurately classify diverse bird species.

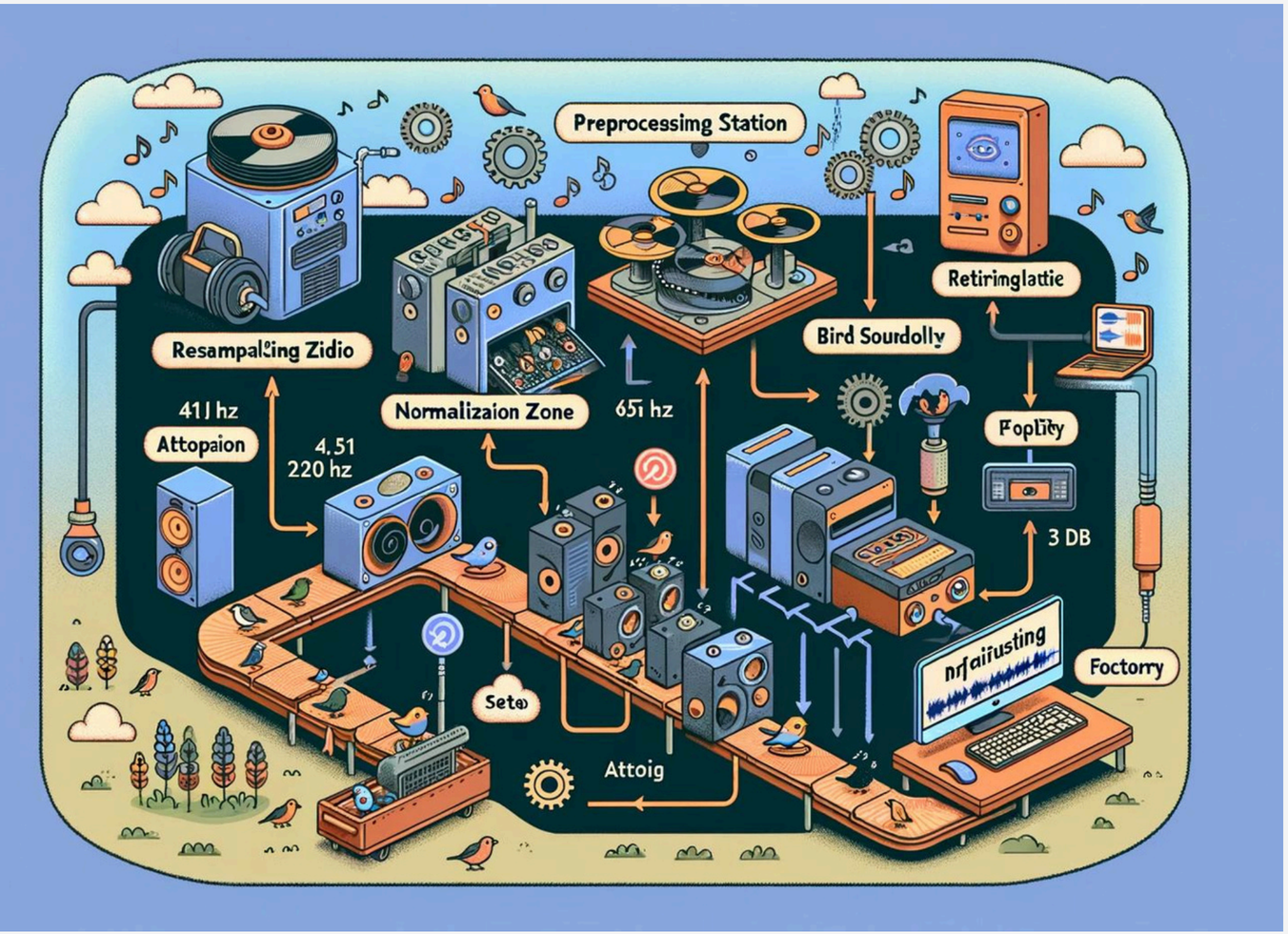


Recent Advancements Identifying Birds by Sound

- Integration of CNNs, NMF, and data augmentation enhances sound classification.
- Improves accuracy in environmental sound classification, including bird sounds.
- Models become more adept at navigating real-world audio complexities.

Dataset Information

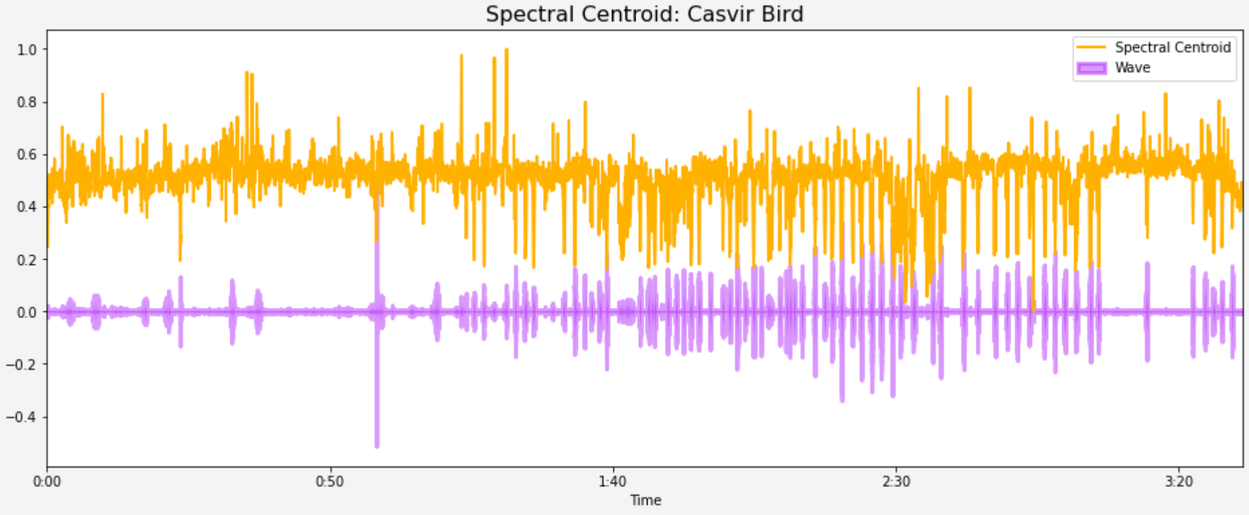
We use Xeno-canto's diverse data to identify bird species by calls.
Xeno-canto is a community-driven platform with extensive bird audio recordings.
Resource offers varied bird sounds in duration, quality, and format.

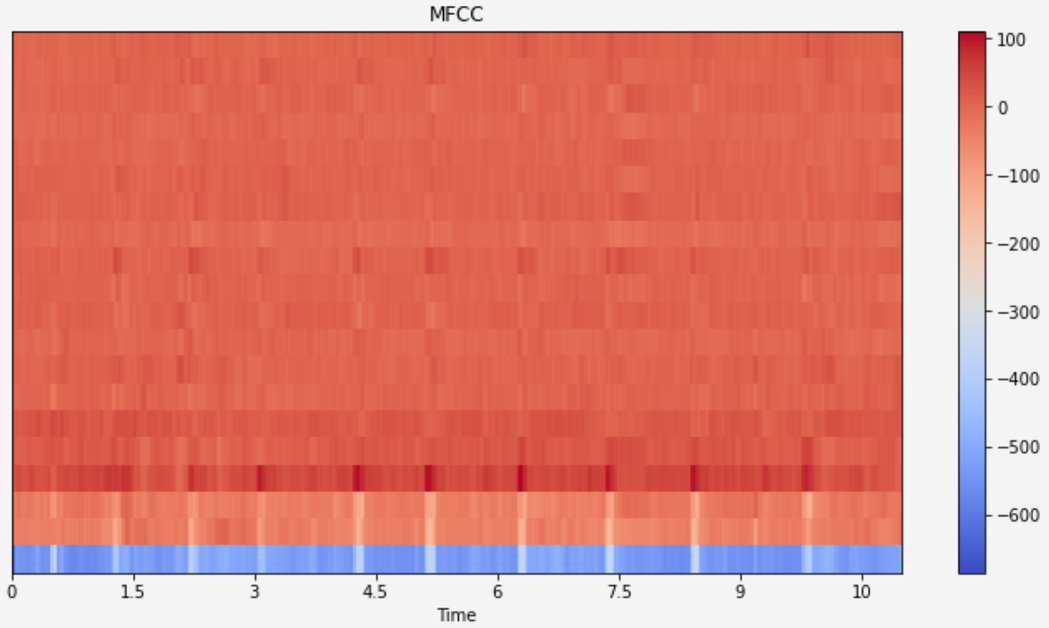


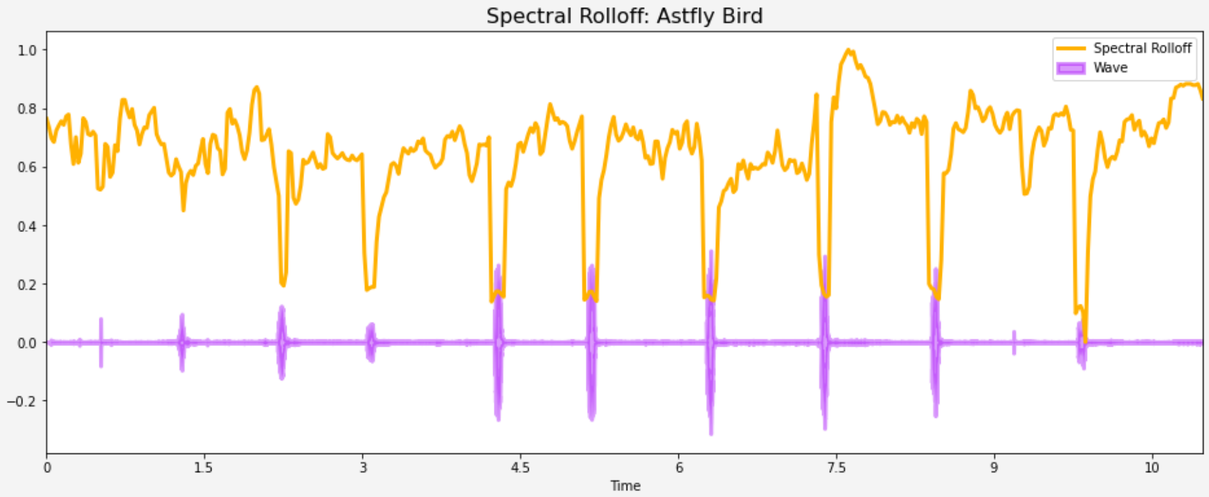
Optimizing Bird Sound Data for Enhanced Model Training

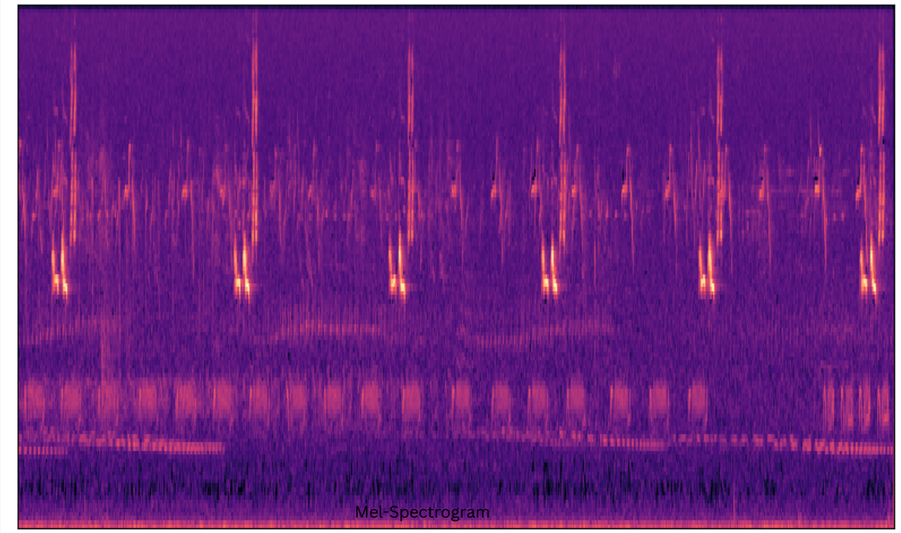
Audio data from Xeno-canto is standardized and segmented into distinct categories like BirdsOnly and NoiseOnly, using techniques such as resampling and filtering, to improve the model's learning efficiency.

Spectrum Related Features









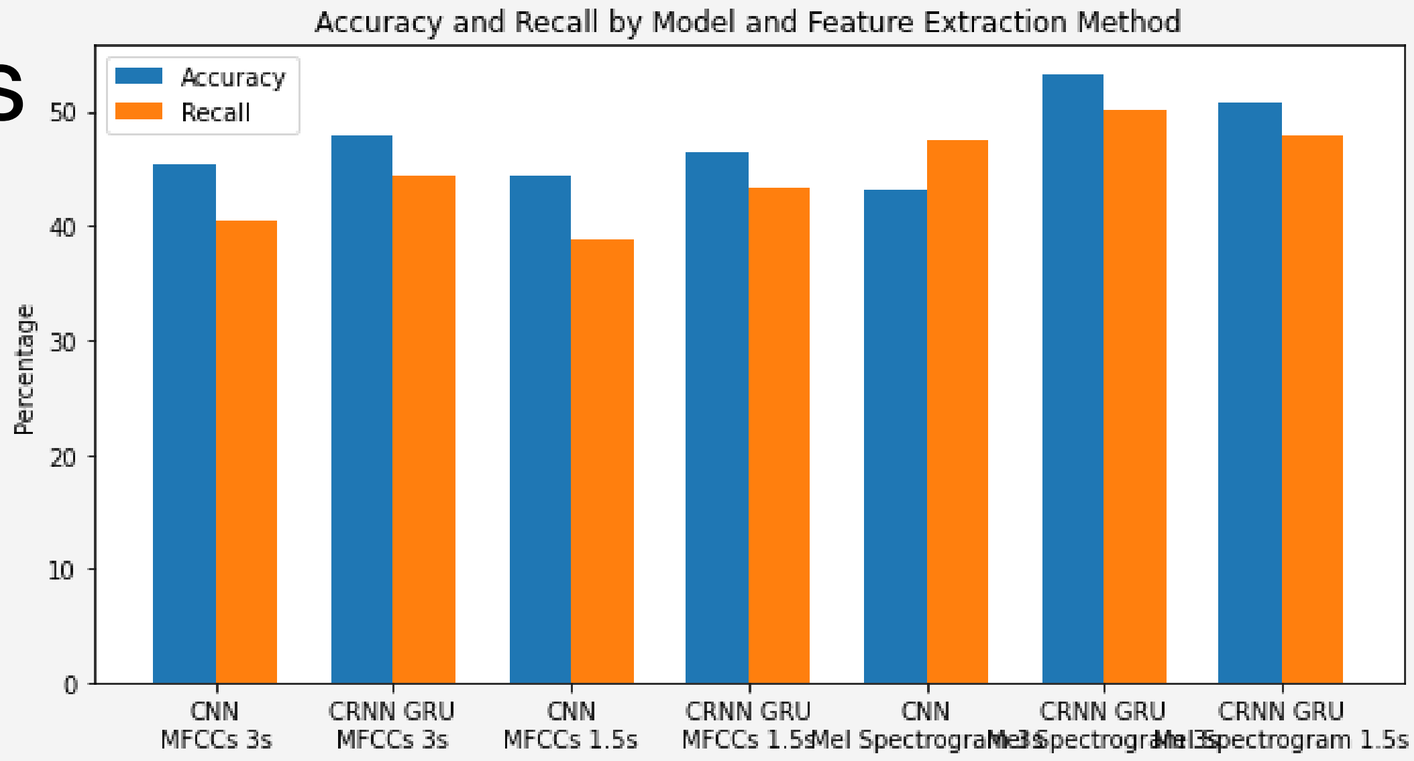
Advanced Architectures for Bird Sound Classification

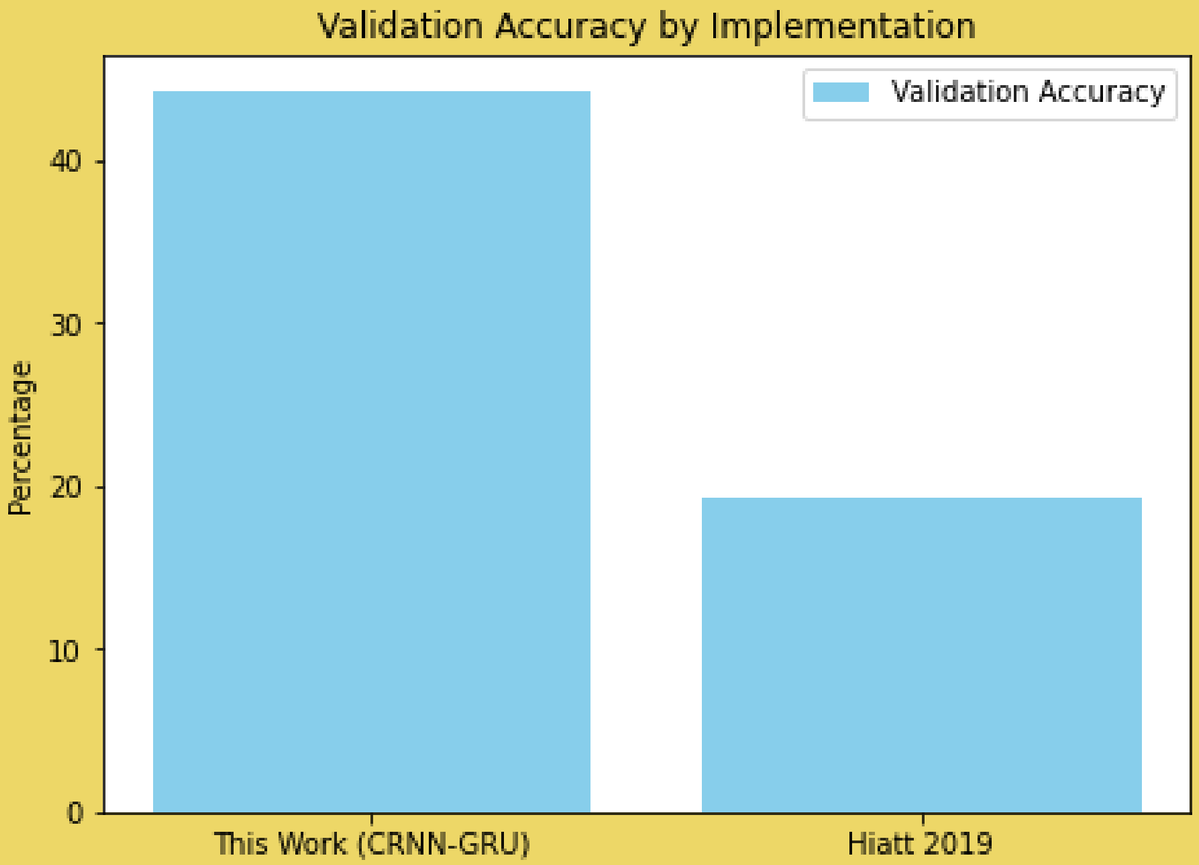
Developed two cutting-edge models leveraging deep learning techniques for the task of bird sound classification. Each model utilizes a unique architecture designed to capture the complex patterns in bird calls.

Feature	CNN	CRNN-GRU
Core Architecture	Convolutional Layers	Convolutional + GRU Layers
Depth of Convolutional Layers	6 Layers	4 Layers
Temporal Feature Processing	Not Applicable	GRU Layers
Dropout Rate	50%	50%
Final Activation	Softmax	Softmax
Total Trainable Parameters	933,211	15,065,915

Evaluating Model Performance: Metrics and Experimental Results

Utilizing an 80/20 training-validation holdout, we tested CNN and CRNN-GRU models' accuracy and recall on MFCC and Mel Spectrogram features for bird sound classification.





Conclusion: Optimized Audio Preprocessing and Deep Learning Success

The optimized CRNN-GRU approach significantly outperformed previous models, achieving a validation accuracy of 44.26% against Hiatt's 19.27%.

```
audio, sample_rate = librosa.load("/kaggle/input/birdclef-2023/train_audio/afghor1/XC156639.ogg")
sample_rate, wav_data = ensure_sample_rate(audio, sample_rate)
Audio(wav_data, rate=sample_rate)
```

[9]:

0:00 / 0:37

[10]:

```
fixed_tm = frame_audio(wav_data)
logits, embeddings = model.infer_tf(fixed_tm[:1])
probabilities = tf.nn.softmax(logits)
argmax = np.argmax(probabilities)
print(f"The audio is from the class {classes[argmax]} (element:{argmax} in the label.csv file), with probability of {p
```

```
The audio is from the class afghor1 (element:46 in the label.csv file), with probability of 0.5590327382087708
```

References

t.ly/dfB7l

THANK YOU

Open House Question