

Klasifikasi Spesies Suara Burung Menggunakan YAMNet dan Random Forest untuk Konservasi Alam

Syaeful Machfud*, Simon Simarmata, Nur Rofiq

Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Pamulang, Tangerang Selatan, Indonesia

Email: ¹*dosen02836@unpam.ac.id, ²dosen02300@unpam.ac.id, ³dosen00376@unpam.ac.id

Email Penulis Korespondensi: dosen02836@unpam.ac.id

Submitted 27-12-2025; Accepted 02-02-2026; Published 28-02-2026

Abstrak

Pemantauan dan identifikasi spesies burung merupakan aspek penting dalam konservasi keanekaragaman hayati, namun metode identifikasi manual berbasis pengamatan langsung dan pendengaran ahli masih memiliki keterbatasan dari segi waktu, biaya, serta subjektivitas. Tantangan lain muncul akibat variasi kualitas rekaman suara, keberadaan noise lingkungan, serta kemiripan pola vokalisasi antarspesies yang menyulitkan proses klasifikasi otomatis. Penelitian ini bertujuan mengembangkan sistem klasifikasi otomatis spesies burung berbasis sinyal akustik dengan mengombinasikan model *YouTube Audio Event Network* (YAMNet) dan algoritma *Random Forest*. YAMNet dimanfaatkan untuk mengekstraksi fitur log Mel spektral yang merepresentasikan karakteristik frekuensi dan temporal suara burung, sedangkan Random Forest digunakan sebagai pengklasifikasi untuk menentukan spesies berdasarkan fitur tersebut. Dataset yang digunakan adalah *Sound of 114 Species of Birds till 2022*, yang mencakup variasi spesies, durasi rekaman, dan kondisi akustik yang kompleks. Hasil penelitian menunjukkan bahwa fitur yang dihasilkan YAMNet mampu membentuk pemisahan antarspesies secara visual melalui *Principal Component Analysis* (PCA), meskipun masih terdapat tumpang tindih pada spesies dengan karakteristik vokalisasi yang serupa. Evaluasi menggunakan *confusion matrix* menunjukkan bahwa beberapa spesies dapat diklasifikasikan dengan tingkat ketepatan yang tinggi, sementara kesalahan klasifikasi terutama terjadi pada spesies dengan pola frekuensi yang mirip. Analisis *Receiver Operating Characteristic* (ROC) menghasilkan nilai *Area Under Curve* (AUC) hingga 0,98 pada spesies tertentu, yang menandakan kemampuan diskriminatif model yang sangat baik. Temuan ini menunjukkan bahwa integrasi YAMNet dan Random Forest berpotensi menjadi solusi yang efisien dan andal untuk mendukung sistem identifikasi spesies burung otomatis dalam konservasi alam.

Kata Kunci: Klasifikasi Suara Burung; YAMNet; Random Forest; Confusion Matrix; Konservasi Alam

Abstract

Monitoring and identification of bird species is an important aspect of biodiversity conservation, but manual identification methods based on direct observation and expert listening still have limitations in terms of time, cost, and subjectivity. Other challenges arise due to variations in the quality of sound recordings, the presence of environmental noise, and the similarity in vocalization patterns between species that make it difficult to automate the classification process. This study aims to develop an automatic classification system of bird species based on acoustic signals by combining the *YouTube Audio Event Network* (YAMNet) model and the *Random Forest* algorithm. YAMNet is utilized to extract spectral log-Mel features that represent the frequency and temporal characteristics of bird sounds, while Random Forest is used as a classifier to determine species based on those features. The dataset used is the *Sound of 114 Species of Birds till 2022*, which includes species variation, recording duration, and complex acoustic conditions. The results showed that the features produced by YAMNet were able to form separation between species visually through *Principal Component Analysis* (PCA), although there was still overlap in species with similar vocalization characteristics. Evaluation using the *confusion matrix* shows that some species can be classified with a high degree of accuracy, while misclassification occurs mainly in species with similar frequency patterns. Receiver Operating Characteristic (ROC) analysis yields *Area Under Curve* (AUC) values of up to 0.98 in certain species, indicating the model's excellent discriminating ability. These findings suggest that the integration of YAMNet and Random Forest has the potential to be an efficient and reliable solution to support automated bird species identification systems in nature conservation.

Keywords: Classification of Bird Sounds; YAMNet; Random Forest; Confusion Matrix; Nature Conservation

1. PENDAHULUAN

Pemantauan dan identifikasi spesies burung merupakan komponen penting dalam upaya konservasi keanekaragaman hayati, karena burung sering digunakan sebagai indikator kesehatan ekosistem dan perubahan lingkungan. Perubahan populasi burung dapat mencerminkan degradasi habitat, perubahan iklim, maupun aktivitas manusia yang berdampak pada keseimbangan ekologi [1][2]. Namun, proses identifikasi spesies burung secara konvensional masih banyak bergantung pada pengamatan visual dan pendengaran oleh ahli ornitologi, yang memerlukan pengalaman tinggi, waktu yang lama, serta biaya yang tidak sedikit. Selain itu, metode manual bersifat subjektif dan sulit diterapkan secara berkelanjutan pada area konservasi yang luas dan terpencil. Kondisi tersebut mendorong perlunya pendekatan otomatis berbasis teknologi yang mampu melakukan identifikasi spesies burung secara efisien, objektif, dan skalabel.

Salah satu pendekatan yang semakin banyak digunakan dalam pemantauan satwa liar adalah analisis suara atau bioacoustic monitoring. Suara burung memiliki karakteristik frekuensi dan temporal yang khas bagi setiap spesies, sehingga dapat dimanfaatkan sebagai dasar klasifikasi [3]. Meskipun demikian, pengembangan sistem klasifikasi suara burung secara otomatis menghadapi berbagai tantangan, antara lain variasi kualitas rekaman, perbedaan durasi audio, keberadaan noise lingkungan, serta kemiripan pola vokalisasi antarspesies. Tantangan ini menjadi semakin kompleks ketika sistem dihadapkan pada dataset berskala besar dengan jumlah spesies yang banyak dan distribusi data yang tidak seimbang, seperti yang umum ditemukan pada rekaman lapangan [4].

Dalam beberapa tahun terakhir, kemajuan di bidang *machine learning* dan *deep learning* telah membuka peluang besar untuk mengatasi permasalahan tersebut. Model berbasis *Convolutional Neural Network* (CNN) dan *transfer learning* terbukti efektif dalam mengekstraksi fitur akustik yang kompleks dari sinyal audio [5][6]. Penelitian oleh [7] menunjukkan bahwa penggunaan CNN *pra latih* seperti VGG16 dan ResNet mampu mencapai akurasi tinggi dalam klasifikasi suara burung, namun performanya sangat bergantung pada kualitas dan jumlah data pelatihan. Selanjutnya, [8] mengusulkan pendekatan semi *supervised learning* untuk mengatasi keterbatasan data berlabel, tetapi model yang digunakan memiliki kompleksitas tinggi dan membutuhkan sumber daya komputasi besar.

Penelitian [9] menggunakan *Fully Convolutional Network* (FCN) untuk klasifikasi multi-spesies burung dan melaporkan hasil yang cukup baik pada dataset berskala menengah. Meskipun demikian, pendekatan tersebut masih terbatas pada jumlah spesies tertentu dan belum sepenuhnya diuji pada data dengan variasi kondisi rekaman yang ekstrem. Di sisi lain, [10] memanfaatkan fitur *log Mel* dan *Gammatone* untuk meningkatkan representasi akustik suara burung, namun penelitian tersebut dilakukan pada lingkungan yang relatif terkendali, sehingga generalisasi model ke kondisi lapangan masih menjadi pertanyaan. Penelitian lainnya oleh [11] memperkenalkan model *AudioProtoPNet* yang bersifat *interpretable*, namun kompleksitas arsitekturnya menyulitkan penerapan pada sistem monitoring *real time* di lapangan.

Berdasarkan tinjauan penelitian terkait tersebut, dapat disimpulkan bahwa sebagian besar penelitian sebelumnya masih menghadapi beberapa keterbatasan utama, yaitu cakupan jumlah spesies yang relatif terbatas, ketergantungan pada model *deep learning* yang kompleks dan mahal secara komputasi, serta kurangnya fokus pada efisiensi dan kemudahan implementasi di lingkungan konservasi nyata. Selain itu, masih sedikit penelitian yang menggabungkan model *deep feature extractor* dengan algoritma *machine learning* yang lebih ringan dan *interpretable* untuk klasifikasi suara burung berskala besar. Kondisi ini menunjukkan adanya *research gap* dalam pengembangan sistem klasifikasi suara burung yang mampu menangani banyak spesies, memiliki performa yang baik, namun tetap efisien dan praktis untuk diterapkan [12].

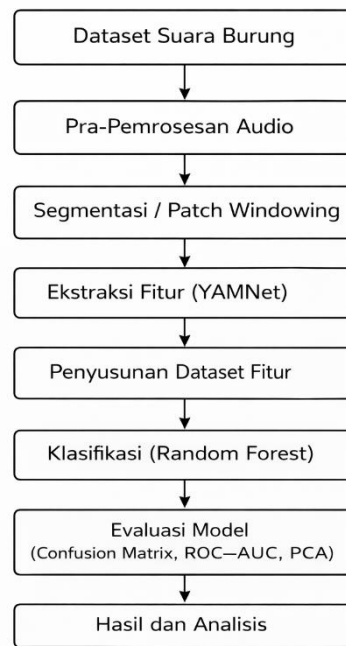
Untuk menjawab celah penelitian tersebut, maka penelitian ini mengusulkan pendekatan klasifikasi otomatis spesies burung dengan menggabungkan model *YouTube Audio Event Network* (YAMNet) dan algoritma *Random Forest*. *YAMNet* dipilih karena merupakan model *transfer learning* yang telah dilatih pada dataset AudioSet berskala besar, sehingga mampu mengekstraksi fitur *log Mel spektral* yang kaya dan representatif tanpa memerlukan pelatihan dari awal [13][14]. Sementara itu, *Random Forest* digunakan sebagai pengklasifikasi karena memiliki keunggulan dalam menangani data berdimensi tinggi, lebih tahan terhadap *overfitting*, serta relatif efisien secara komputasi dibandingkan model *deep learning end to end*.

Penelitian ini bertujuan untuk mengembangkan dan mengevaluasi sistem klasifikasi otomatis spesies burung berbasis suara menggunakan kombinasi *YAMNet* dan *Random Forest* dengan memanfaatkan dataset *Sound of 114 Species of Birds till 2022*. Tujuan khusus dari penelitian ini adalah menganalisis kemampuan ekstraksi fitur *YAMNet* dalam memisahkan karakteristik suara antarspesies, mengevaluasi performa klasifikasi menggunakan *confusion matrix* dan *Receiver Operating Characteristic* (ROC), serta menilai potensi penerapan sistem dalam konteks konservasi alam. Diharapkan hasil penelitian ini dapat memberikan kontribusi ilmiah dalam pengembangan sistem pemantauan keanekaragaman hayati berbasis akustik yang akurat, efisien, dan aplikatif, serta menjadi dasar bagi penelitian lanjutan dalam pengembangan sistem monitoring satwa liar cerdas di masa depan.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahapan penelitian disusun secara berurutan mulai dari pengumpulan data, pra pemrosesan audio, ekstraksi fitur, proses klasifikasi, hingga evaluasi kinerja model. Alur metodologi ini bertujuan memastikan bahwa sistem yang dihasilkan mampu mengklasifikasi spesies burung secara akurat, efisien dan dapat diterapkan pada konservasi alam. Adapun tahapan metode penelitian adalah sebagai berikut:



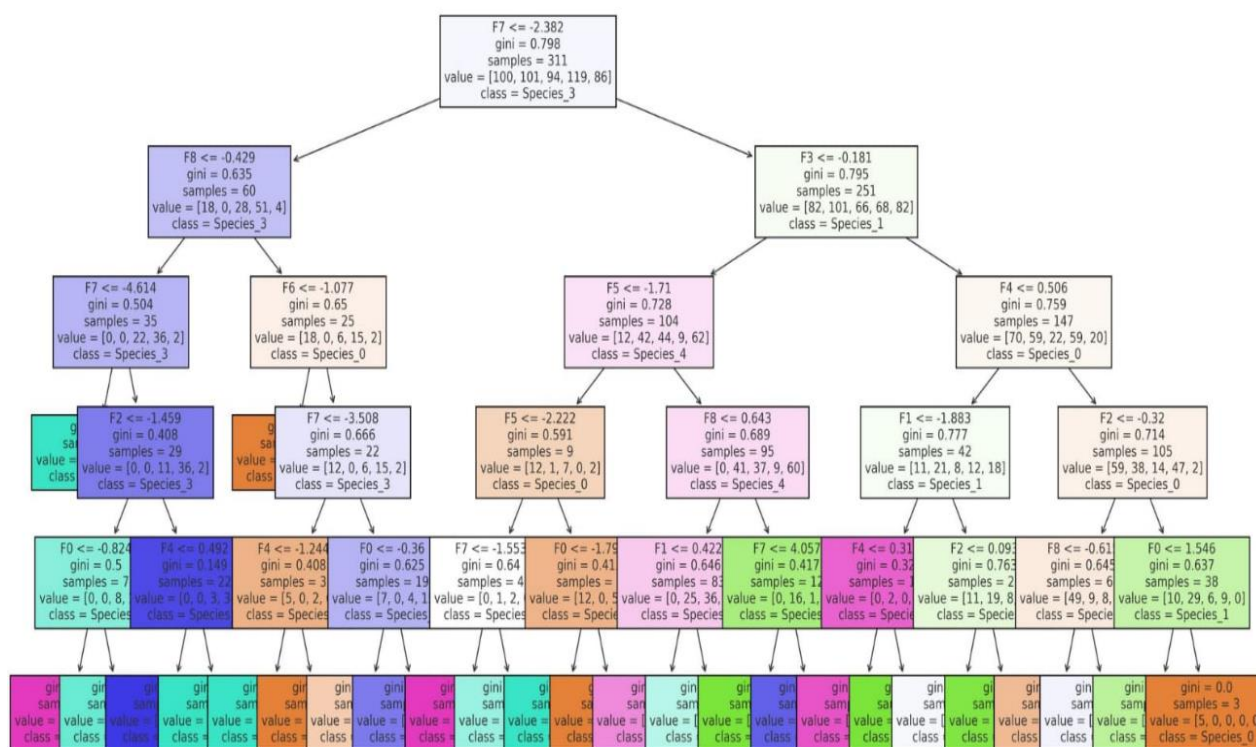
Gambar 1. Metodologi Penelitian

Berdasarkan tahapan penelitian pada gambar 1, dapat dijelaskan masing – masing tahapan sebagai berikut:

- Dataset Suara Burung : sumber data pada penelitian ini diambil dari kaggle dengan pemilihan *dataset Sound of 114 species of Bird till 2022*. Dimana file ini berisi rekaman suara burung dari 114 *species* berbeda dengan variasi jenis vokalisasi, durasi rekaman, dan kondisi lingkungan.
- Pra Pemrosesan Audio : pada tahap ini, seluruh file audio diseragamkan ke format mono dengan laju *sampling* 16 kHz agar sesuai dengan kebutuhan model *YAMNet*. Dimana proses ini bertujuan untuk menghilangkan perbedaan format audio dan menjaga konsistensi data [15]. Selain itu, normalisasi amplitudo dilakukan untuk mengurangi pengaruh perbedaan tingkat volume antar rekaman, sehingga karakteristik suara burung dapat direpresentasikan secara lebih stabil.
- Segmentasi / *Patch Windowing* : Audio yang telah dipra-proses selanjutnya dibagi menjadi segmen-segmen pendek menggunakan teknik *patch windowing*. Segmentasi dilakukan dengan panjang jendela 0,96 detik dan hop size 0,48 detik, mengikuti konfigurasi standar *YAMNet* [16]. Pendekatan ini memungkinkan sistem menangkap informasi temporal suara burung secara lebih detail dan meningkatkan jumlah sampel data dari setiap rekaman, sehingga variasi vokalisasi dapat dipelajari dengan lebih baik.
- Ekstraksi Fitur (*YAMNet*) : Setiap segmen audio diekstraksi menggunakan model *YAMNet* yang telah dilatih sebelumnya pada dataset *AudioSet*. *YAMNet* menghasilkan fitur *log-Mel spectrogram* dan *embedding* akustik berdimensi tinggi yang merepresentasikan pola frekuensi dan temporal suara burung. Penggunaan *transfer learning* pada tahap ini memungkinkan ekstraksi fitur yang kaya dan informatif tanpa perlu melatih model dari awal, sehingga lebih efisien secara komputasi [17].
- Penyusunan Dataset Fitur : Fitur hasil ekstraksi *YAMNet* kemudian digabungkan dengan label spesies burung untuk membentuk dataset fitur terstruktur. Setiap baris data merepresentasikan satu segmen audio dengan vektor fitur dan label spesies yang sesuai. Dataset ini selanjutnya dibagi menjadi data pelatihan dan data pengujian untuk proses klasifikasi dan evaluasi model.
- Klasifikasi (*Random Forest*) : Tahap klasifikasi dilakukan menggunakan algoritma *Random Forest* yang memanfaatkan kumpulan pohon keputusan untuk memprediksi spesies burung berdasarkan fitur *YAMNet*. *Random Forest* dipilih karena mampu menangani data berdimensi tinggi, relatif tahan terhadap *overfitting*, serta memiliki performa yang stabil pada dataset dengan banyak kelas seperti pada penelitian ini. *Random forest* merupakan algoritma pembelajaran ensemble yang menggabungkan sejumlah pohon keputusan yang bekerja sama bersama sama untuk meningkatkan akurasi tugas klasifikasi maupun *regresi* [18]. Menghitung entropi dan gini dengan menggunakan rumus berikut:

$$\text{Entropy} = \sum_i^C p_i * \log_2 p_i \quad (1)$$

$$\text{Gini} = 1 - \sum_i^C p_i^2 \quad (2)$$



Gambar 2. Pohon Keputusan Model *Random Forest*

Gambar 2 menggambarkan salah satu pohon keputusan (*decision tree*) dalam model *Random Forest* yang digunakan untuk mengklasifikasikan suara burung berdasarkan fitur akustik hasil ekstraksi dari model YAMNet menggunakan dataset *Sound of 114 Species of Birds till 2022*. Proses dimulai dari node akar (*root node*) di bagian atas dengan kondisi $F7 \leq -2.382$, di mana seluruh sampel suara pertama kali dibagi menjadi dua kelompok besar berdasarkan nilai fitur F7. Node ini memiliki nilai Gini = 0.798, yang berarti tingkat heterogenitas masih tinggi karena berisi campuran dari beberapa spesies (*Species_0* hingga *Species_4*). Dari node akar, pohon bercabang ke kiri dan kanan, dimana cabang kiri mewakili kondisi True (nilai F7 memenuhi syarat ≤ -2.382) dan cabang kanan untuk False (nilai $F7 > -2.382$). Setiap percabangan berikutnya menggunakan fitur lain seperti F8, F6, F3, F4, dan F5 untuk mempersempit keputusan klasifikasi hingga setiap node daun (*leaf node*) di bagian bawah hanya berisi satu kelas spesies dominan. Nilai Gini pada setiap node menunjukkan seberapa bersih pemisahan kelas di titik tersebut; semakin kecil nilainya (mendekati 0), semakin murni node tersebut, artinya sampel di dalamnya hampir seluruhnya berasal dari satu spesies. Sementara itu, nilai *samples* menunjukkan jumlah data suara yang masuk ke node tersebut, dan *value* = [a, b, c, d, e] menggambarkan jumlah sampel dari masing-masing spesies yang ada dalam node. Warna pada setiap kotak node juga memiliki makna penting: setiap warna merepresentasikan kelas atau spesies burung dominan yang diprediksi oleh model pada node tersebut. Misalnya, warna ungu menunjukkan *Species_3*, warna hijau untuk *Species_0*, warna jingga untuk *Species_4*, dan warna biru muda untuk *Species_1*. Semakin pekat warna pada node, semakin besar tingkat keyakinan model bahwa data pada node tersebut termasuk ke dalam spesies tersebut.

- g. Evaluasi Model : Kinerja model dievaluasi menggunakan confusion matrix untuk menganalisis tingkat ketepatan dan kesalahan klasifikasi pada setiap spesies. Selain itu, kurva ROC dan nilai AUC digunakan untuk menilai kemampuan diskriminatif model, sedangkan *Principal Component Analysis* (PCA) dimanfaatkan untuk memvisualisasikan separabilitas fitur antarspesies secara dua dimensi. *Confusion Matrix* digunakan sebagai alat utama untuk mengevaluasi performa klasifikasi multikelas dengan membandingkan label spesies sebenarnya (*true label*) dan label hasil prediksi model (*predicted label*). Pada penelitian ini, baris pada *confusion matrix* merepresentasikan spesies burung yang sebenarnya, sedangkan kolom menunjukkan spesies hasil prediksi *Random Forest*. *Principal Component Analysis* (PCA) digunakan sebagai metode reduksi dimensi untuk memvisualisasikan distribusi embedding hasil ekstraksi YAMNet yang berdimensi tinggi. Dengan mereduksi fitur *embedding* ke dalam dua komponen utama (PC1 dan PC2), PCA memungkinkan pengamatan pola sebaran dan pengelompokan segmen suara burung dalam ruang fitur dua dimensi. [19][20].
- h. Hasil dan Analisis : Tahap akhir penelitian adalah analisis hasil klasifikasi untuk menilai efektivitas metode yang diusulkan. Hasil evaluasi digunakan untuk mengidentifikasi spesies yang dapat diklasifikasikan dengan baik serta spesies yang masih mengalami tumpang tindih vokalisasi. Analisis ini menjadi dasar untuk menarik kesimpulan dan memberikan rekomendasi pengembangan lanjutan sistem klasifikasi suara burung otomatis dalam konteks konservasi alam.

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Dataset yang digunakan adalah *Sound of 114 Species of Birds till 2022* dari Kaggle, yang berisi 2.161 file audio dalam format .wav yang merepresentasikan 114 spesies burung berbeda. Setiap file berisi suara kicauan atau panggilan spesifik dari masing-masing spesies. Data ini dikumpulkan dari berbagai sumber global melalui platform *Xeno Canto*, yang dikenal menyediakan rekaman suara burung dari seluruh dunia. Pada tahap ini dilakukan pula proses data preparation seperti normalisasi format audio ke mono 16 kHz, pemotongan durasi agar seragam (misalnya 3 sampai 5 detik), dan pemberian label sesuai nama spesies. Langkah ini memastikan bahwa data yang masuk ke model berada dalam kondisi bersih, konsisten, dan siap untuk ekstraksi fitur.

Tabel 1. Dataset Suara Burung

Spesies Burung	Jumlah Rekaman	Durasi rata – rata (detik)	Negara Asal	Jenis Suara
Common Ostrich (<i>Struthio camelus australis</i>)	56	45.2	South Africa	Call
Small billed tinamous (<i>Crypturellus parvirostris</i>)	38	32.8	Brazil	Song
Andean Guan (<i>Penelope Montagnii</i>)	41	29.4	Peru	Call
Band tailed guan (<i>penelope argyrotis</i>)	28	37.5	Colombia	Song
White browed guan (<i>penelope jacucaca</i>)	34	30.1	Argentina	Song
.....				
Total (114 Spesies)	2.161 file audio	34.6 detik/file	17 Negara	Call dan Song

3.2 Pra Pemrosesan Data

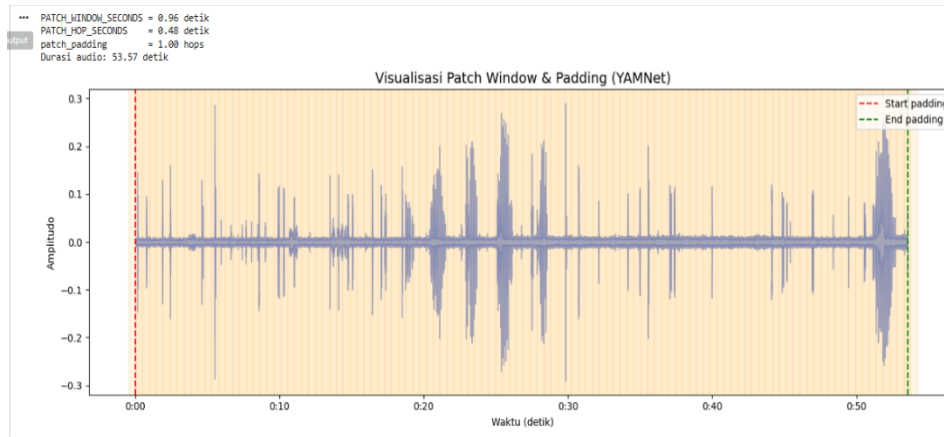
Pra pemrosesan dilakukan untuk menyeragamkan format audio agar sesuai dengan kebutuhan model *YAMNet* dan meningkatkan kualitas fitur yang diekstraksi. Dataset awal memiliki variasi format, durasi, dan setelah pra pemrosesan, seluruh audio berada dalam kondisi yang konsisten. Berikut adalah hasil dari pra pemrosesan yang sudah dilakukan:

Tabel 2. Hasil Pra Pemrosesan dan Ekstraksi Fitur

Tahapan Pra Pemrosesan	Kondisi Awal Dataset	Hasil setelah pra pemrosesan
Format Audio	Beragam (MP3/WAV)	WAV
Channel	Mono dan Stereo	Mono
Sampling rate	Bervariasi (8 kHz – 44.1 kHz)	16 kHz
Normalisasi Amplitudo	Tidak seragam	Rentang -1 sampai 1
Durasi Audio	Bervariasi (detik – menit)	Tetap (proses dengan <i>patch windowing</i>)
Noise Lingkungan	Ada (alami)	Dipertahankan
Audio siap ekstraksi fitur	Tidak	Ya (sesuai input <i>YAMNet</i>)

3.3 Visualisasi Patch Window dan Padding (YAMNet)

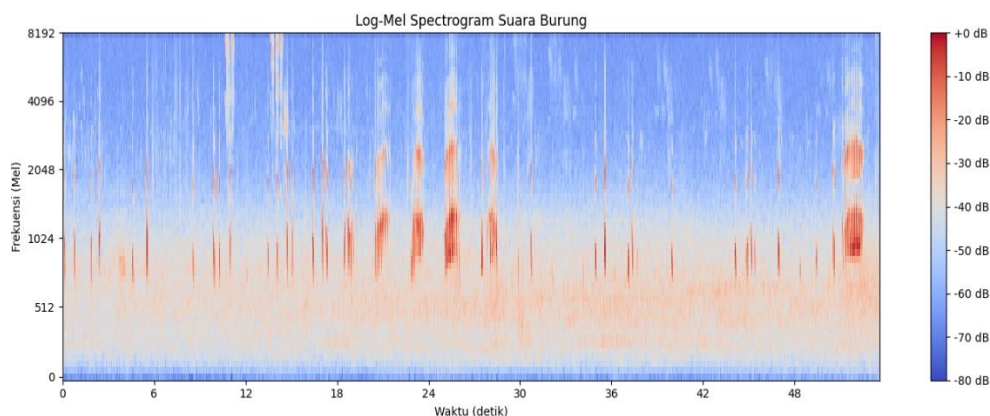
Tahapan Visualisasi *Patch Window* dan *Padding* (*YAMNet*) menunjukkan proses segmentasi sinyal audio burung sebelum dilakukan ekstraksi fitur menggunakan model *YAMNet*. Sinyal audio berdurasi sekitar 53,57 detik dibagi menjadi potongan-potongan waktu (*patch*) dengan panjang 0,96 detik dan jarak antar *patch* (*hop*) sebesar 0,48 detik, sehingga terjadi tumpang tindih (*overlap*) antar *patch* sebesar 50%. Teknik ini bertujuan untuk menjaga kesinambungan informasi temporal suara burung yang bersifat dinamis dan tidak stasioner. Area berwarna latar pada grafik merepresentasikan pembagian patch secara berurutan sepanjang sinyal audio, sementara garis putus-putus merah dan hijau masing-masing menandai padding awal dan padding akhir yang diterapkan untuk memastikan seluruh bagian sinyal audio dapat diproses secara konsisten oleh *YAMNet*. Padding ini penting karena panjang rekaman audio sering kali tidak habis terbagi secara tepat oleh ukuran patch, sehingga tanpa padding sebagian informasi di awal atau akhir sinyal dapat terabaikan. Dengan pendekatan *patch windowing* dan *padding* ini, setiap variasi vokalisasi burung baik berupa panggilan singkat, jeda, maupun perubahan intensitas suara yang dapat direpresentasikan secara merata ke dalam patch, sehingga meningkatkan kemampuan *YAMNet* dalam mengekstraksi fitur akustik yang stabil dan informatif untuk tahap klasifikasi selanjutnya.



Gambar 3. Visualisasi Patch Window dan Padding (YAMNet)

3.4 Tahap Representasi Akustik

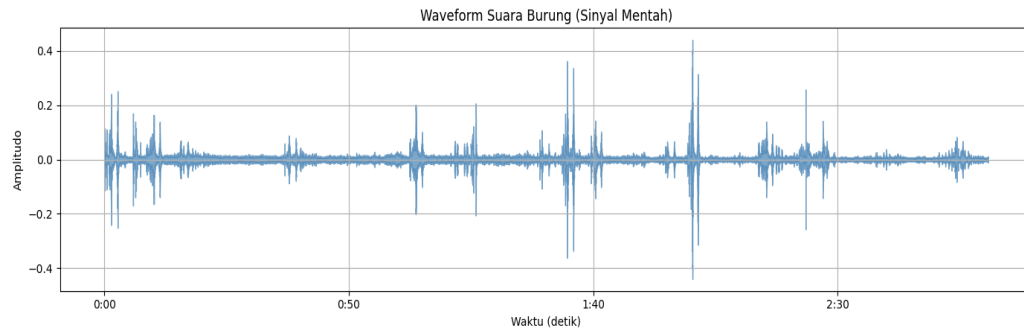
Representasi akustik diperoleh dalam bentuk log Mel spectrogram, yang memetakan energi sinyal audio pada skala frekuensi Mel terhadap waktu. Hasil yang ditampilkan oleh log-Mel spectrogram dari suara burung *Small billed* Tinamou (*Crypturellus parvirostris*) yang digunakan sebagai representasi akustik awal sebelum ekstraksi fitur oleh model *YAMNet*. Sumbu horizontal merepresentasikan waktu (dalam detik), sedangkan sumbu vertikal menunjukkan frekuensi pada skala Mel, yang dirancang untuk meniru persepsi pendengaran manusia terhadap perbedaan frekuensi. Intensitas warna pada spectrogram menggambarkan energi sinyal dalam satuan desibel (dB), di mana warna yang lebih hangat (merah oranye) menunjukkan energi suara yang lebih tinggi dan warna yang lebih dingin (biru) menunjukkan energi yang lebih rendah. Pola energi yang muncul secara periodik pada rentang frekuensi menengah hingga tinggi mencerminkan karakteristik vokalisasi khas spesies ini, seperti panggilan berulang dengan struktur temporal yang konsisten. *Representasi log Mel spectrogram* ini sangat penting karena mampu menangkap informasi frekuensi dan temporal secara simultan, sekaligus mereduksi pengaruh noise latar yang umum terdapat pada rekaman lapangan. Dengan demikian, spectrogram ini menjadi input yang efektif bagi *YAMNet* dalam mengekstraksi embedding akustik yang diskriminatif untuk membedakan spesies burung pada tahap klasifikasi selanjutnya.



Gambar 4. Log Mel Spectrogram

3.5 Tahap Visualisasi Sinyal Mentah (Waveform Audio)

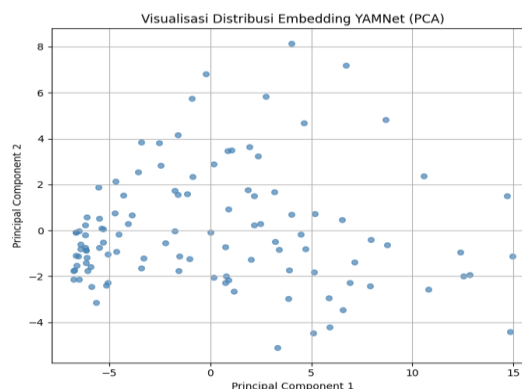
Hasil tahap visualisasi sinyal mentah (*raw audio*) dari dataset *Sound of 114 Species of Birds till 2022* ini memperlihatkan representasi suara burung dalam domain waktu, dengan sumbu horizontal menunjukkan durasi rekaman (dalam detik) dan sumbu vertikal menunjukkan amplitudo sinyal audio. Terlihat adanya variasi amplitudo yang cukup signifikan sepanjang durasi rekaman, yang menandakan kemunculan beberapa segmen vokalisasi burung dengan intensitas tinggi, diselingi oleh bagian-bagian dengan amplitudo rendah yang merepresentasikan jeda atau latar belakang lingkungan. Nilai pada sumbu waktu menunjukkan bahwa durasi rekaman berada pada kisaran lebih dari dua menit, yang menegaskan bahwa dataset ini memuat rekaman dengan durasi cukup panjang sehingga mampu menangkap beberapa pola vokalisasi dalam satu file audio. Sementara itu, nilai amplitudo sinyal berada pada rentang sekitar $-0,45$ hingga $+0,45$, yang mencerminkan variasi energi suara selama proses perekaman. Segmen dengan amplitudo mendekati nol menunjukkan bagian rekaman yang relatif hening atau didominasi oleh noise lingkungan, sedangkan lonjakan amplitudo yang tajam baik pada nilai positif maupun negatif yang menandakan kemunculan suara burung dengan intensitas tinggi, seperti kicauan atau panggilan spesifik.



Gambar 5. Waveform Audio

3.6 Tahap Ekstraksi Fitur (*Distibusi Patch Embedding/PCA*)

Visualisasi Distribusi *Embedding YAMNet* (PCA) menunjukkan hasil pemetaan fitur akustik yang diekstraksi dari dataset *Sound of 114 Species of Birds till 2022* ke dalam ruang dua dimensi menggunakan *Principal Component Analysis* (PCA). Nilai pada sumbu horizontal, yaitu *Principal Component 1* (PC1), merepresentasikan komponen utama pertama hasil analisis *Principal Component Analysis* (PCA). PC1 menangkap variasi terbesar dari seluruh fitur *embedding YAMNet* yang diekstraksi dari segmen audio burung. Secara konseptual, nilai PC1 mencerminkan perbedaan dominan dalam karakteristik akustik, seperti variasi energi spektral, rentang frekuensi utama, dan intensitas vokalisasi. Titik dengan nilai PC1 yang semakin besar atau kecil menunjukkan segmen suara burung yang memiliki karakteristik akustik yang semakin berbeda dibandingkan segmen lainnya. Sementara itu, nilai pada sumbu vertikal, yaitu *Principal Component 2* (PC2), merepresentasikan komponen utama kedua yang menangkap variasi terbesar berikutnya setelah PC1, namun tetap bersifat independen (ortogonal) terhadap PC1. PC2 umumnya menggambarkan variasi akustik tambahan, seperti perbedaan pola temporal, ritme vokalisasi, atau distribusi frekuensi menengah dan tinggi. Nilai PC2 yang berbeda menunjukkan adanya variasi sekunder dalam pola suara burung yang tidak sepenuhnya dijelaskan oleh PC1. Dengan demikian, kombinasi nilai PC1 dan PC2 membentuk posisi setiap titik pada grafik, yang secara visual menggambarkan hubungan kemiripan dan perbedaan antarsegmen suara burung. Semakin dekat posisi titik-titik tersebut, semakin mirip karakteristik akustiknya, sedangkan jarak yang semakin jauh menunjukkan perbedaan pola vokalisasi yang lebih signifikan antarsegmen atau antarspesies.

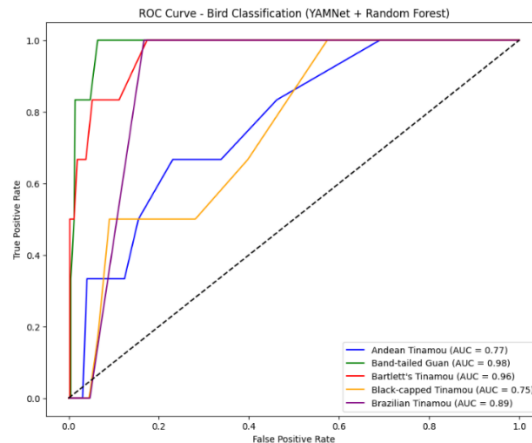


Gambar 6. Distribusi Patch Embedding

3.7 Klasifikasi (*ROC Curve*)

ROC Curve Bird Classification (YAMNet + Random Forest) menunjukkan hasil evaluasi kinerja model klasifikasi spesies burung yang dikembangkan dalam penelitian ini dengan menggunakan dataset *Sound of 114 Species of Birds till 2022*. Setiap kurva ROC merepresentasikan kemampuan model dalam membedakan satu spesies burung tertentu terhadap spesies lainnya berdasarkan fitur akustik hasil ekstraksi *YAMNet* yang kemudian diklasifikasikan menggunakan *Random Forest*. Sumbu horizontal menunjukkan *False Positive Rate* (FPR), sedangkan sumbu vertikal menunjukkan *True Positive Rate* (TPR), yang menggambarkan *trade off* antara tingkat kesalahan klasifikasi dan tingkat keberhasilan deteksi spesies. Hasil menunjukkan bahwa model memiliki performa yang bervariasi antarspesies. Spesies *Band tailed Guan* menunjukkan kinerja terbaik dengan nilai *Area Under Curve* (AUC) sebesar 0.98, yang menandakan kemampuan diskriminatif yang sangat tinggi dalam membedakan spesies tersebut dari kelas lainnya. Spesies *Bartlett's Tinamou* juga menunjukkan performa yang sangat baik dengan AUC sebesar 0.96, sedangkan *Brazilian Tinamou* memperoleh AUC sebesar 0.89, yang masih tergolong sangat baik. Di sisi lain, *Andean Tinamou* dan *Black capped Tinamou* memiliki nilai AUC yang lebih rendah, masing-masing 0.77 dan 0.75, yang mengindikasikan adanya kesulitan model dalam membedakan spesies-spesies ini akibat kemiripan pola vokalisasi dengan spesies lain dalam dataset. Secara keseluruhan, kurva ROC ini menunjukkan bahwa integrasi *YAMNet* dan *Random Forest* mampu menghasilkan performa klasifikasi

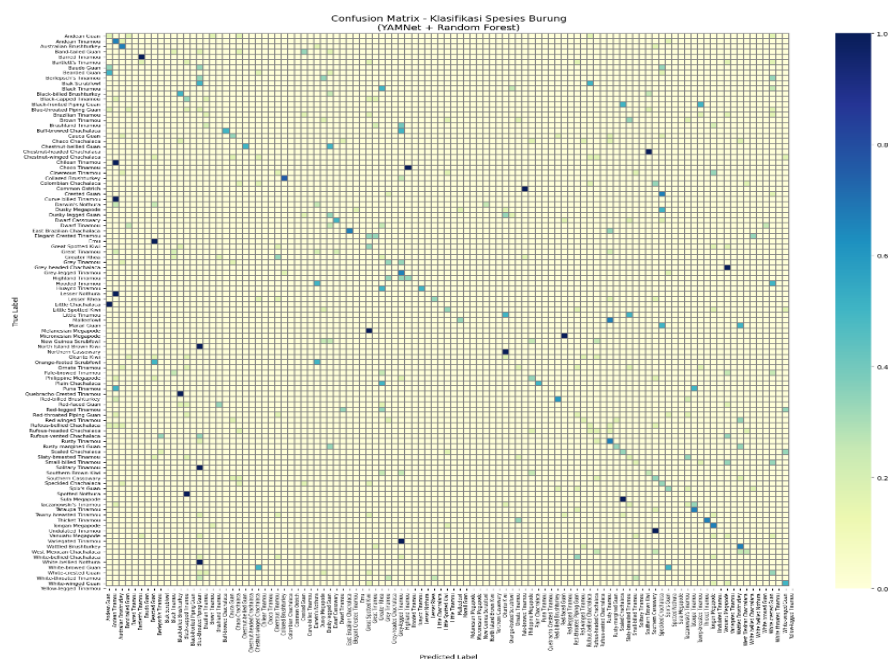
yang kuat untuk sebagian besar spesies, sekaligus mengungkap adanya tantangan pada spesies dengan karakteristik suara yang saling tumpang tindih.



Gambar 7. ROC Curve

3.8 Evaluasi (Confusion Matrix)

Confusion matrix hasil klasifikasi multikelas spesies burung disini terdapat sumbu vertikal (*True Label*) merepresentasikan spesies burung yang sebenarnya, sedangkan sumbu horizontal (*Predicted Label*) menunjukkan spesies hasil prediksi model. Intensitas warna pada setiap sel menyatakan proporsi prediksi, di mana warna yang semakin gelap menunjukkan frekuensi atau proporsi prediksi yang lebih tinggi. Hasil yang terlihat menunjukkan bahwa sebagian prediksi benar terkonsentrasi pada diagonal utama, yang menandakan bahwa model mampu mengklasifikasikan sejumlah spesies dengan tepat. Ini mengindikasikan bahwa fitur akustik hasil ekstraksi *YAMNet* cukup representatif untuk membedakan pola vokalisasi spesies tertentu. Namun demikian, diagonal yang tidak sepenuhnya dominan dan banyaknya titik berwarna di luar diagonal utama menunjukkan adanya kesalahan klasifikasi antarspesies. Kesalahan ini terutama terjadi pada spesies-spesies yang memiliki kemiripan karakteristik vokalisasi, baik dari segi frekuensi dominan, ritme, maupun durasi panggilan, sehingga fitur akustiknya saling tumpang tindih. Selain itu, penyebaran nilai yang relatif jarang dan tipis di banyak sel mencerminkan tingginya kompleksitas klasifikasi multikelas dengan jumlah spesies yang besar (114 spesies) serta kemungkinan ketidakseimbangan jumlah sampel antarspesies dalam dataset. Secara keseluruhan, *confusion matrix* ini menunjukkan bahwa model *YAMNet + Random Forest* telah mampu mengenali pola suara burung secara umum, namun performanya masih bervariasi antarspesies. Hasil ini menegaskan bahwa pendekatan yang digunakan sudah menjanjikan untuk identifikasi otomatis spesies burung, sekaligus membuka peluang pengembangan lanjutan melalui penyeimbangan data, optimasi parameter model, atau integrasi metode klasifikasi yang lebih adaptif untuk meningkatkan akurasi pada spesies dengan pola vokalisasi yang mirip.



Gambar 8. Confusion Matrix

4. KESIMPULAN

Kesimpulan dari penelitian ini menunjukkan bahwa sistem klasifikasi spesies burung berbasis sinyal akustik menggunakan kombinasi *YouTube Audio Event Network (YAMNet)* dan *Random Forest* mampu menjawab permasalahan utama dalam identifikasi spesies burung secara otomatis yang selama ini masih bergantung pada metode manual yang memerlukan waktu, biaya, dan keahlian khusus. Hasil analisis sinyal mentah pada dataset memperlihatkan kompleksitas data audio yang tinggi, ditandai dengan variasi amplitudo, durasi, serta keberadaan *noise* lingkungan, sehingga menuntut metode ekstraksi fitur yang robust. Penerapan *YAMNet* terbukti efektif dalam mengekstraksi representasi akustik yang informatif, sebagaimana ditunjukkan oleh visualisasi PCA yang memperlihatkan pemisahan antarsegmen suara, meskipun masih terdapat tumpang tindih pada spesies dengan karakteristik vokalisasi yang serupa. Evaluasi kinerja melalui *ROC Curve* menunjukkan nilai *Area Under Curve (AUC)* yang tinggi pada beberapa spesies, mencapai hingga 0.98, yang menandakan kemampuan diskriminatif model yang sangat baik. *Confusion matrix* mengindikasikan bahwa model mampu mengklasifikasikan sebagian spesies dengan tingkat ketepatan yang memadai, meskipun performa belum merata akibat jumlah kelas yang besar dan ketidakseimbangan data. Keterbatasan penelitian ini terletak pada variasi jumlah sampel antarspesies dan kemiripan pola vokalisasi, sehingga penelitian selanjutnya disarankan untuk mengeksplorasi penyeimbangan data dan metode klasifikasi yang lebih adaptif guna meningkatkan akurasi dan generalisasi sistem.

REFERENCES

- [1] F. Terranova *et al.*, "Windy events detection in big bioacoustics datasets using a pre-trained Convolutional Neural Network," *Sci. Total Environ.*, vol. 949, p. 174868, 2024, doi: <https://doi.org/10.1016/j.scitotenv.2024.174868>.
- [2] P. Lauha *et al.*, "Bird Sounds Global - model builder : An end-to-end workflow for building locally fine-tuned bird classifiers," pp. 1–20.
- [3] B. Ghani, T. Denton, S. Kahl, and H. Klinck, "Global birdsong embeddings enable superior transfer learning for bioacoustic classification," *Sci. Rep.*, vol. 13, no. 1, p. 22876, 2023, doi: 10.1038/s41598-023-49989-z.
- [4] I. C. P, D. K. R, and M. R, "Bird Sound Identification System using Deep Learning," *Procedia Comput. Sci.*, vol. 233, pp. 597–603, 2024, doi: <https://doi.org/10.1016/j.procs.2024.03.249>.
- [5] S. Kahl, C. M. Wood, M. Eibl, and H. Klinck, "BirdNET: A deep learning solution for avian diversity monitoring," *Ecol. Inform.*, vol. 61, p. 101236, 2021, doi: <https://doi.org/10.1016/j.ecoinf.2021.101236>.
- [6] R. D. Beason, R. Riesch, and J. Koricheva, "Investigating the effects of tree species diversity and relative density on bird species richness with acoustic indices," *Ecol. Indic.*, vol. 154, p. 110652, 2023, doi: <https://doi.org/10.1016/j.ecolind.2023.110652>.
- [7] J. Xie and M. Zhu, "Acoustic Classification of Bird Species Using an Early Fusion of Deep Features," *Birds*, vol. 4, no. 1, pp. 138–147, 2023, doi: 10.3390/birds4010011.
- [8] P. N. Andono and Y. Sari, "Feature Selection on Gammatone Cepstral Coefficients for Bird Voice Classification Using Particle Swarm Optimization," vol. 16, no. 1, pp. 254–264, 2023, doi: 10.22266/ijies2023.0228.23.
- [9] M. K. Baowaly *et al.*, "Deep transfer learning-based bird species classification using mel spectrogram images," *PLoS One*, vol. 19, no. 8, p. e0305708, Aug. 2024, [Online]. Available: <https://doi.org/10.1371/journal.pone.0305708>
- [10] D. Trivedi, R. Sarmukaddam, and V. C. Gandhi, "Deep Learning for Urban Sound Classification: Using CNN and YAMNet Model Integration BT - Advanced Technologies in Electronics, Communications and Signal Processing," K. K. Koganti, S. R. E., and N. Gupta, Eds., Cham: Springer Nature Switzerland, 2026, pp. 332–349.
- [11] B. Williams *et al.*, "Using tropical reef, bird and unrelated sounds for superior transfer learning in marine bioacoustics," *Philos. Trans. R. Soc. B Biol. Sci.*, vol. 380, no. 1928, p. 20240280, 2025, doi: 10.1098/rstb.2024.0280.
- [12] S. D. H. Permana, G. Saputra, B. Arifitama, Yaddarabullah, W. Caesarendra, and R. Rahim, "Classification of bird sounds as an early warning method of forest fires using Convolutional Neural Network (CNN) algorithm," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 7, pp. 4345–4357, 2022, doi: <https://doi.org/10.1016/j.jksuci.2021.04.013>.
- [13] K. K. Mohammed, E. I. A. El-Latif, N. E. El-Sayad, A. Darwish, and A. E. Hassanien, "Radio frequency fingerprint-based drone identification and classification using Mel spectrograms and pre-trained YAMNet neural," *Internet of Things*, vol. 23, p. 100879, 2023, doi: <https://doi.org/10.1016/j.iot.2023.100879>.
- [14] T. Tuncer, E. Akbal, and S. Dogan, "Multileveled ternary pattern and iterative ReliefF based bird sound classification," *Appl. Acoust.*, vol. 176, p. 107866, 2021, doi: <https://doi.org/10.1016/j.apacoust.2020.107866>.
- [15] A. Revathi and N. Sasikaladevi, "Robust sound-based bird classification using multiple features and random forest classifier," *Int. J. Speech Technol.*, vol. 28, no. 1, pp. 117–127, 2025, doi: 10.1007/s10772-025-10168-2.
- [16] Y.-C. Wei, W.-L. Chen, M.-N. Tuanmu, S.-S. Lu, and M.-T. Shiao, "Advanced montane bird monitoring using self-supervised learning and transformer on passive acoustic data," *Ecol. Inform.*, vol. 84, p. 102927, 2024, doi: <https://doi.org/10.1016/j.ecoinf.2024.102927>.
- [17] L. Liu *et al.*, "YAMNet-based transfer learning for compact noise classification in urban and wireless systems," *EURASIP J. Wirel. Commun. Netw.*, vol. 2025, no. 1, p. 74, 2025, doi: 10.1186/s13638-025-02483-8.
- [18] S. Amini, M. Saber, H. Rabiei-Dastjerdi, and S. Homayouni, "Urban Land Use and Land Cover Change Analysis Using Random Forest Classification of Landsat Time Series," *Remote Sens.*, vol. 14, no. 11, 2022, doi: 10.3390/rs14112654.
- [19] R. Pahuja and A. Kumar, "Sound-spectrogram based automatic bird species recognition using MLP classifier," *Appl. Acoust.*, vol. 180, p. 108077, 2021, doi: <https://doi.org/10.1016/j.apacoust.2021.108077>.
- [20] F. J. Bravo Sanchez, M. R. Hossain, N. B. English, and S. T. Moore, "Bioacoustic classification of avian calls from raw sound waveforms with an open-source deep learning architecture," *Sci. Rep.*, vol. 11, no. 1, p. 15733, 2021, doi: 10.1038/s41598-021-95076-6.