

Analisis Kemampuan Model *Hybrid CNN-GRU* dalam Memahami Dinamika Temporal Suara Burung Berbasis *Audio Spectrogram*

¹Mario Lusiano Klau, ^{2*}Ryan Putranda Kristianto

^{1,2} Program Studi Ilmu Informatika, Fakultas Teknik, Universitas Katolik Darma Cendika

Penulis Korespondensi: marioklau86@gmail.com, ryan@ukdc.ac.id

Abstract. *Audio spectrogram-based bird sound classification typically relies on Convolutional Neural Networks (CNNs) to extract spatial features; however, this approach often fails to optimally utilize the temporal sequence information inherent in audio signals, despite the complex temporal characteristics of bird vocalization patterns. This study aims to analyze the capability of a hybrid Convolutional Neural Network-Gated Recurrent Unit (CNN-GRU) model to capture the temporal dynamics of bird sounds. A quantitative experimental approach was employed using a dataset of 850 audio recordings from five Indonesian urban bird species sourced from Xeno-Canto; this dataset was expanded to 1,000 samples via data augmentation to address class imbalance. The audio underwent normalization, resampling, and noise reduction before being converted into mel-spectrograms—using the Short-Time Fourier Transform (STFT)—to serve as model input. The model's ability to comprehend temporal dynamics was evaluated through the visualization of GRU hidden states using t-SNE and a Frame Shuffle Test. Test results demonstrated an overall accuracy of 76%, with precision, recall, and F1-score values of 77%, 76%, and 76%, respectively, on the test data. Visualization of the GRU hidden states revealed that the model's internal representations evolved progressively in line with the frame sequence, whereas shuffling the frame order resulted in reduced confidence scores and altered class predictions. These findings confirm that the CNN-GRU model does not merely rely on spatial features from the spectrogram but actively leverages inter-frame temporal relationships during classification, demonstrating that integrating a GRU into the hybrid architecture effectively captures the temporal dynamics of bird sounds.*

Keywords: *audio spectrogram; temporal dynamics; Gated Recurrent Unit (GRU); hybrid CNN-GRU; bird sound classification*

Abstrak. Klasifikasi suara burung berbasis *audio spectrogram* umumnya mengandalkan Convolutional Neural Network (CNN) untuk mengekstraksi fitur spasial, namun pendekatan tersebut belum optimal dalam memanfaatkan informasi urutan waktu yang terkandung pada sinyal audio, padahal pola vokalisasi burung memiliki karakteristik temporal yang kompleks. Penelitian ini bertujuan menganalisis kemampuan model *hybrid* Convolutional Neural Network-Gated Recurrent Unit (CNN-GRU) dalam memahami dinamika temporal suara burung. Penelitian menggunakan pendekatan kuantitatif dengan metode eksperimen pada dataset 850 rekaman suara dari lima spesies burung urban Indonesia yang diperoleh dari Xeno-Canto, kemudian ditambah melalui augmentasi data menjadi 1.000 sampel untuk mengatasi ketidakseimbangan kelas. Audio diolah melalui normalisasi, resampling, penghilangan noise, dan dikonversi menjadi mel-spectrogram menggunakan *Short-Time Fourier Transform* (STFT) sebagai masukan model. Kemampuan model dalam memahami dinamika temporal dianalisis melalui visualisasi hidden state GRU menggunakan t-SNE serta pengujian Frame Shuffle Test. Hasil pengujian menunjukkan model memperoleh akurasi 76%, dengan precision, recall, dan F1-score masing-masing sebesar 77%, 76%, dan 76% pada data uji. Visualisasi hidden state GRU menunjukkan representasi internal model berubah secara bertahap mengikuti urutan frame, sedangkan pengacakan urutan frame menyebabkan penurunan nilai confidence dan perubahan kelas prediksi. Temuan ini membuktikan bahwa model CNN-GRU tidak hanya mengandalkan fitur spasial dari spectrogram, tetapi juga secara aktif memanfaatkan hubungan temporal antarframe dalam proses klasifikasi, sehingga integrasi GRU pada arsitektur hybrid terbukti mampu menangkap dinamika temporal suara burung secara efektif.

Kata kunci: *audio spectrogram; dinamika temporal; Gated Recurrent Unit (GRU); hybrid CNN-GRU; klasifikasi suara burung*

1. LATAR BELAKANG

Indonesia merupakan salah satu negara dengan tingkat keanekaragaman burung yang tinggi. Berdasarkan data Status Burung di Indonesia tahun 2025 yang diterbitkan oleh Perhimpunan Pelestarian Burung Liar Indonesia (BirdLife Indonesia Association), tercatat 1.835 spesies burung yang terklasifikasi dalam 24 ordo dan 129 famili, dengan Passeriformes mendominasi sebesar 52% dari total spesies (Alberto & Hermanto, 2023). Identifikasi spesies burung yang akurat sangat penting untuk mendukung pemantauan populasi, konservasi keanekaragaman hayati, serta pengambilan kebijakan lingkungan (Hernández-Navarro et al., 2024). Salah satu pendekatan yang dapat digunakan adalah identifikasi berdasarkan suara burung, karena metode berbasis audio memungkinkan pemantauan dilakukan secara non-invasif tanpa mengganggu habitat alami burung (Chang & Sinnott, 2022). Sinyal audio burung umumnya direpresentasikan dalam bentuk spectrogram, yaitu representasi visual dari perubahan frekuensi terhadap waktu ((Foggia et al., 2023); (Levy et al., 2022)), sehingga data audio dapat diproses menyerupai citra dan mempermudah ekstraksi fitur menggunakan pendekatan Convolutional Neural Network (CNN).

Meskipun demikian, klasifikasi suara burung tetap menjadi tantangan karena sinyal audio memiliki karakteristik temporal yang kompleks, seperti durasi sinyal, ritme, struktur sekuens, dan interval antar-vokalisasi, sehingga proses klasifikasi tidak cukup hanya mengandalkan ekstraksi fitur spasial (Baowaly et al., 2024). CNN telah banyak diterapkan pada klasifikasi audio dan terbukti efektif mengenali pola spasial pada spectrogram ((Ferdiana et al., 2021); (Ozcan & Gungor, 2025)), namun arsitektur ini memiliki keterbatasan dalam menangkap hubungan antarwaktu karena fokusnya lebih diarahkan pada fitur lokal (Myagila et al., 2025), dan berpotensi mengalami overfitting pada dataset yang terbatas (Revadekar et al., 2023). Keterbatasan tersebut mendorong penggunaan arsitektur berbasis Recurrent Neural Network (RNN) untuk melengkapi CNN dalam memodelkan dependensi temporal. Gated Recurrent Unit (GRU) dipilih sebagai komponen temporal karena memiliki struktur gerbang yang lebih sederhana dibandingkan Long Short-Term Memory (LSTM), sehingga kompleksitas komputasinya lebih ringan, namun tetap mampu mempelajari dependensi temporal pada data sekuensial ((Alsyabani et al., 2022); (Elsayed et al., 2019); (Zarzycki & Ławryńczuk, 2021)).

Berdasarkan uraian tersebut, artikel ini menyajikan bagian dari hasil penelitian yang mengembangkan model hybrid CNN-GRU untuk klasifikasi suara burung, dengan fokus pembahasan pada permasalahan: bagaimana kemampuan model hybrid CNN-GRU dalam memahami dinamika temporal suara burung? Tujuan penelitian ini adalah menganalisis dan mengevaluasi kemampuan model hybrid CNN-GRU dalam menangkap dan memanfaatkan pola dinamika temporal pada spectrogram suara burung, melalui kombinasi evaluasi performa klasifikasi serta analisis mekanistik terhadap representasi internal model. Hasil penelitian ini diharapkan dapat memberikan kontribusi empiris mengenai peran komponen rekuren pada arsitektur hybrid dalam tugas klasifikasi bioakustik, sekaligus menjadi dasar bagi pengembangan sistem pemantauan burung berbasis suara yang lebih akurat.

2. METODE PENELITIAN

Pendekatan dan Dataset Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen untuk menguji kemampuan model *hybrid* CNN-GRU dalam mengklasifikasikan suara burung sekaligus memahami dinamika temporalnya. Dataset yang digunakan terdiri atas 850 rekaman suara dari lima spesies burung urban Indonesia yang diperoleh dari platform Xeno-Canto, yaitu Cucak Kutilang (*Pycnonotus aurigaster*) 168 data, Burung Gereja (*Passer montanus*) 200 data, Tekukur Biasa (*Spilopelia chinensis*) 200 data, Kipasan Belang (*Rhipidura javanica*) 146 data, dan Cabak Kota (*Caprimulgus affinis*) 136 data. Karena distribusi data antar kelas tidak seimbang, dilakukan oversampling melalui augmentasi data berupa penambahan noise, pergeseran waktu, dan perubahan kecepatan pada kelas minoritas, sehingga jumlah data setiap kelas disamakan menjadi 200 sampel (total 1.000 sampel). Dataset kemudian dibagi menjadi data pelatihan dan validasi dengan proporsi 70:30, sedangkan data pengujian sebanyak 200 sampel diambil dari sumber terpisah, yaitu rekaman media sosial, untuk memastikan evaluasi dilakukan pada data yang benar-benar baru.

Praproses Data

Seluruh berkas audio di-resampling ke frekuensi 16.000 Hz dan diseragamkan durasinya menjadi 10 detik melalui *zero-padding* atau pemotongan. Amplitudo audio

dinormalisasi menggunakan peak normalization ke rentang $[-1, 1]$, dan noise latar dihilangkan menggunakan metode spectral gating dari pustaka noisereduce. Setiap sinyal audio selanjutnya dikonversi menjadi mel-spectrogram berukuran 128×128 piksel dengan tiga kanal menggunakan Short-Time Fourier Transform (STFT), dengan parameter *window size* 1024, *hop length* 256, N-FFT 1024, dan 128 mel-band. Parameter praproses disajikan pada Tabel 1.

Parameter	Nilai	Keterangan
Sample rate	16.000 Hz	Resampling dari berbagai sampling rate asli
Durasi audio	10 detik	Padding/pemotongan
Window size (STFT)	1024	Pembuatan spectrogram
Hop length	256	Overlap antar-window
N-FFT	1024	Sama dengan window size
Jumlah mel-band	128	Resolusi frekuensi pada mel-spectrogram
Normalisasi	Peak normalization	Amplitudo dinormalisasi ke rentang $[-1, 1]$
Pengurangan noise	Spectral gating	Menggunakan pustaka noisereduce

Tabel 1. Parameter Praproses Audio

Arsitektur Model Hybrid CNN-GRU

Model hybrid CNN-GRU diimplementasikan menggunakan framework TensorFlow. Bagian CNN terdiri atas tiga blok konvolusi (Conv2D-BatchNormalization-ReLU-MaxPooling2D) dengan jumlah filter 32, 64, dan 128 secara berurutan, yang berfungsi mengekstraksi fitur spasial hierarkis dari spectrogram. Output dari blok

konvolusi terakhir di-reshape menjadi urutan fitur dan diumpankan ke dua lapisan GRU bertumpuk (128 dan 64 unit) yang bertugas mempelajari dependensi temporal antarframe. Representasi akhir GRU diteruskan ke lapisan dense dengan fungsi aktivasi softmax untuk menghasilkan probabilitas lima kelas burung. Model dilatih menggunakan optimizer Adam (learning rate 0,001), fungsi loss categorical cross-entropy, ukuran batch 16, dan maksimum 50 epoch dengan early stopping.

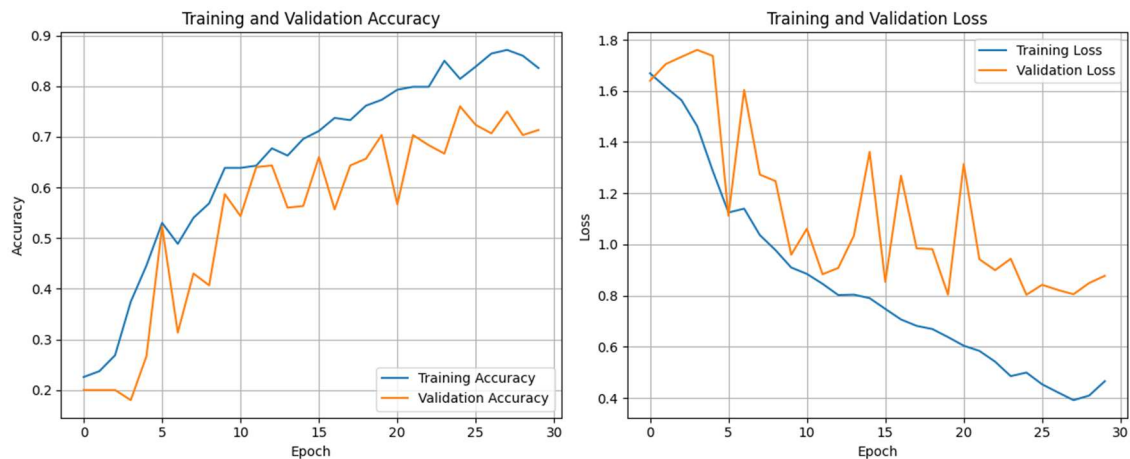
Metode Analisis Dinamika Temporal

Untuk menjawab permasalahan mengenai kemampuan model dalam memahami dinamika temporal, dilakukan dua eksperimen analisis mekanistik. Pertama, visualisasi hidden state GRU pada setiap langkah waktu (frame) direduksi dimensinya menggunakan t-distributed Stochastic Neighbor Embedding (t-SNE) untuk mengamati apakah representasi internal model berubah mengikuti urutan waktu. Kedua, dilakukan Frame Shuffle Test, yaitu pengujian dengan mengacak urutan frame pada spectrogram tanpa mengubah isi setiap frame. Apabila performa dan tingkat keyakinan (confidence) prediksi menurun setelah urutan frame diacak, hal tersebut menjadi bukti empiris bahwa model memanfaatkan hubungan temporal antarframe, bukan hanya pola spasial pada masing-masing frame. Kinerja klasifikasi model secara keseluruhan dievaluasi menggunakan akurasi, precision, recall, F1-score, dan confusion matrix.

3. HASIL DAN PEMBAHASAN

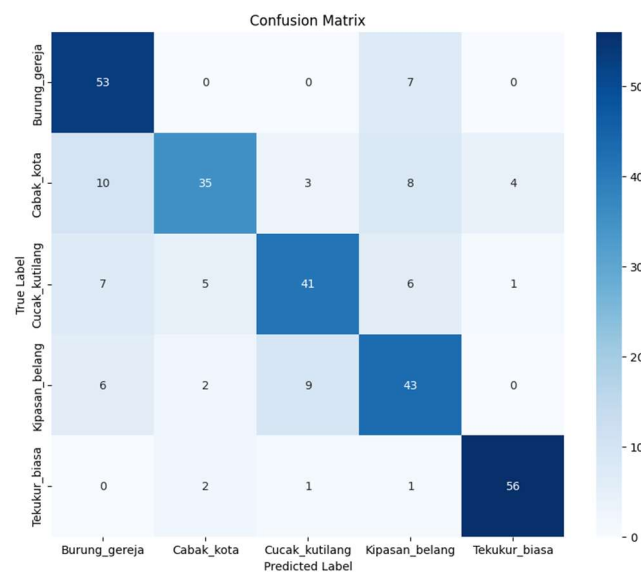
Performa Klasifikasi Model

Proses pelatihan model dilakukan selama 30 epoch efektif dengan early stopping. Gambar 1 menunjukkan bahwa training accuracy meningkat dari sekitar 23% menjadi 87%, sedangkan validation accuracy meningkat dari sekitar 20% menjadi 75%. Training loss menurun dari 1,66 menjadi 0,45, sementara validation loss juga menurun meskipun berfluktuasi pada beberapa epoch. Selisih akurasi pelatihan dan validasi yang berada pada kisaran 10-15% mengindikasikan adanya overfitting ringan, namun model tetap memiliki kemampuan generalisasi yang cukup baik.



Gambar 1. Kurva Akurasi dan Loss Pelatihan Model CNN-GRU

Pada data pengujian sebanyak 300 sampel, model memperoleh akurasi sebesar 76%, dengan nilai precision, recall, dan F1-score rata-rata masing-masing sebesar 0,77, 0,76, dan 0,76 (Tabel 2). Confusion matrix (Gambar 2) menunjukkan bahwa sebagian besar prediksi terkonsentrasi pada diagonal utama. Kelas Tekukur Biasa memperoleh performa terbaik (F1-score 0,93), sedangkan kelas Cabak Kota memperoleh performa terendah (F1-score 0,67). Kesalahan klasifikasi paling banyak terjadi antara kelas Kipasan Belang dan Cucak Kutilang, serta antara Burung Gereja dan Kipasan Belang, yang mengindikasikan adanya kemiripan pola spektral maupun temporal pada pasangan kelas tersebut.



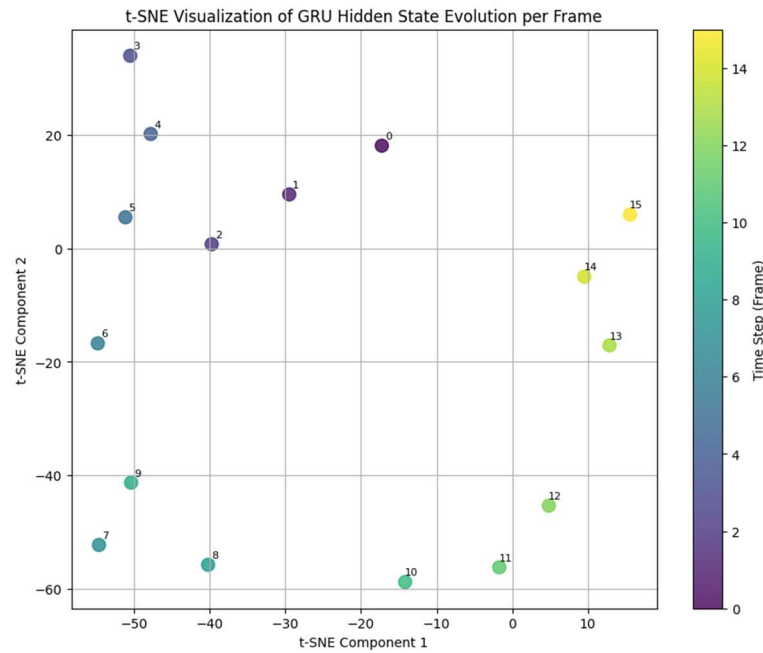
Gambar 2. Confusion Matrix Hasil Pengujian Model

Kelas Burung	Precision	Recall	F1-Score	Support
Burung Gereja	0,70	0,88	0,78	60
Cabak Kota	0,80	0,58	0,67	60
Cucak Kutilang	0,76	0,68	0,72	60
Kipasan Belang	0,66	0,72	0,69	60
Tekukur Biasa	0,92	0,93	0,93	60
Akurasi	-	-	0,76	300
Rata-rata Makro	0,77	0,76	0,76	300
Rata-rata Tertimbang	0,77	0,76	0,76	300

Tabel 2. Hasil Classification Report Model Hybrid CNN-GRU

Visualisasi Hidden State GRU

Gambar 3 menyajikan hasil visualisasi hidden state GRU menggunakan t-SNE untuk satu sampel urutan suara burung sepanjang 16 frame. Warna gradasi ungu hingga kuning merepresentasikan urutan waktu dari frame awal hingga frame akhir. Pada frame awal (frame 0-6), hidden state berada pada sisi kiri grafik. Memasuki frame pertengahan (frame 7-12), posisi hidden state bergeser ke bagian bawah grafik, dan pada frame akhir (frame 13-15) berpindah ke sisi kanan sehingga membentuk pola representasi yang berbeda dari frame-frame sebelumnya. Pergeseran posisi yang konsisten mengikuti urutan waktu ini menunjukkan bahwa representasi internal GRU terus diperbarui sepanjang pemrosesan sekuens, bukan berada pada satu titik yang tetap.



Gambar 3. Visualisasi t-SNE dari Hidden State GRU per Frame

Hasil Frame Shuffle Test

Tabel 3 menyajikan hasil pengujian Frame Shuffle Test pada dua sampel uji. Pada kondisi urutan frame normal, sampel pertama diprediksi sebagai Cabak Kota dengan confidence 0,54, dan sampel kedua diprediksi sebagai Tekukur Biasa dengan confidence 0,99. Setelah urutan frame diacak, prediksi pada kedua sampel berubah: sampel pertama menjadi Tekukur Biasa (confidence 0,42) dan sampel kedua menjadi Cucak Kutilang (confidence 0,42), disertai penurunan confidence pada kedua kasus.

Sampel	Kondisi	Prediksi	Confidence
1	Normal	Cabak Kota	0,54
1	Shuffle	Tekukur Biasa	0,42
2	Normal	Tekukur Biasa	0,99
2	Shuffle	Cucak Kutilang	0,42

Tabel 3. Perbandingan Hasil Frame Shuffle Test

Diskusi/Pembahasan

Hasil pelatihan model menunjukkan bahwa arsitektur hybrid CNN-GRU mampu belajar secara efektif, dengan CNN berperan mengekstraksi pola spasial dari spectrogram dan GRU memanfaatkan urutan fitur tersebut untuk mempelajari hubungan temporal antarframe. Selisih akurasi pelatihan dan validasi yang moderat (10-15%) menunjukkan model belum mengalami overfitting yang berlebihan, sehingga hasil evaluasi selanjutnya dapat dianggap cukup representatif terhadap kemampuan generalisasi model.

Bukti utama yang menjawab permasalahan penelitian mengenai kemampuan model dalam memahami dinamika temporal diperoleh dari dua eksperimen mekanistik. Pertama, visualisasi t-SNE terhadap hidden state GRU memperlihatkan bahwa representasi internal model tidak statis, melainkan bergeser secara sistematis mengikuti perkembangan urutan frame, dari sisi kiri pada frame awal, ke bagian bawah pada frame pertengahan, hingga ke sisi kanan pada frame akhir. Pola pergeseran yang konsisten ini mengindikasikan bahwa GRU mengakumulasi dan memperbarui informasi dari frame-frame sebelumnya, bukan memproses setiap frame secara independen. Temuan ini sejalan dengan prinsip kerja GRU sebagai varian RNN yang dirancang untuk mempertahankan dependensi jangka panjang pada data sekuensial dengan struktur gerbang yang lebih sederhana dibandingkan LSTM.

Kedua, hasil Frame Shuffle Test memberikan bukti kuantitatif yang memperkuat temuan tersebut. Apabila model hanya mengandalkan fitur spasial dari CNN, pengacakan urutan frame semestinya tidak memengaruhi hasil prediksi karena informasi visual pada setiap frame tetap sama. Namun, hasil pengujian menunjukkan perubahan kelas prediksi pada seluruh sampel uji disertai penurunan confidence yang cukup besar, dari 0,54 menjadi 0,42 pada sampel pertama dan dari 0,99 menjadi 0,42 pada sampel kedua. Perubahan ini menunjukkan bahwa hubungan temporal antarframe berkontribusi nyata terhadap keputusan klasifikasi model, sehingga performa model bergantung pada kombinasi fitur spasial dari CNN dan pemodelan temporal dari GRU, bukan pada salah satu komponen saja.

Temuan ini juga relevan dengan hasil classification report, khususnya pada kelas dengan pola temporal yang khas dan konsisten seperti Tekukur Biasa, yang memperoleh

F1-score tertinggi (0,93) karena panggilannya berulang dengan interval waktu yang teratur sehingga lebih mudah dipelajari oleh lapisan GRU. Sebaliknya, kelas Cabak Kota memperoleh F1-score terendah (0,67), yang diduga berkaitan dengan variasi temporal yang lebih tinggi atau kemiripan pola dengan kelas lain, sehingga representasi temporal yang terbentuk kurang terpisah secara jelas antarkelas. Pola kesalahan klasifikasi yang terkonsentrasi pada pasangan kelas dengan karakteristik vokalisasi serupa turut mengindikasikan bahwa kemampuan model dalam memisahkan dinamika temporal antarspesies masih dipengaruhi oleh tingkat kemiripan akustik alami di antara spesies burung tersebut.

Secara keseluruhan, hasil ini memperkuat argumen bahwa keterbatasan CNN dalam menangkap hubungan antarwaktu dapat diatasi melalui integrasi komponen rekuren seperti GRU, tanpa menambah kompleksitas komputasi sebesar LSTM. Dengan demikian, dapat disimpulkan bahwa model hybrid CNN-GRU pada penelitian ini memiliki kemampuan yang terbukti secara empiris dalam memahami dan memanfaatkan dinamika temporal suara burung, baik ditinjau dari perubahan representasi internal (hidden state) maupun dari sensitivitas performa terhadap urutan frame. Meskipun demikian, penelitian ini masih memiliki keterbatasan, antara lain jumlah kelas burung yang terbatas pada lima spesies urban serta indikasi overfitting ringan, sehingga generalisasi model pada spesies dan kondisi akustik yang lebih beragam perlu diuji lebih lanjut pada penelitian selanjutnya.

4. KESIMPULAN DAN SARAN

Berdasarkan hasil evaluasi dan analisis mekanistik yang dilakukan, dapat disimpulkan bahwa model hybrid CNN-GRU terbukti mampu memahami dan memanfaatkan dinamika temporal suara burung. Model memperoleh akurasi klasifikasi sebesar 76% pada data uji, dengan performa terbaik pada kelas Tekukur Biasa (F1-score 0,93) dan performa terendah pada kelas Cabak Kota (F1-score 0,67). Kemampuan model dalam memahami dinamika temporal dibuktikan melalui visualisasi hidden state GRU yang menunjukkan representasi internal model berubah secara bertahap mengikuti urutan frame, serta melalui hasil Frame Shuffle Test yang menunjukkan penurunan confidence dan perubahan kelas prediksi ketika urutan frame diacak. Kedua bukti tersebut secara konsisten menunjukkan bahwa model tidak hanya mengandalkan fitur spasial dari CNN,

tetapi juga secara aktif memanfaatkan hubungan temporal antarframe melalui lapisan GRU dalam proses klasifikasi suara burung.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Program Studi Ilmu Informatika, Fakultas Teknik, Universitas Katolik Darma Cendika, serta dosen pembimbing atas arahan dan bimbingan yang diberikan selama proses penelitian ini

DAFTAR REFERENSI

- Alberto, J., & Hermanto, D. (2023). Klasifikasi Jenis Burung Menggunakan Metode CNN Dan Arsitektur ResNet-50. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 10(3), 34–46.
- Alsyabani, O. M. A., Utami, E., Raharjo, S., & Hartanto, A. D. (2022). Stacked LSTM-GRU Model for Traffic Anomalies Detection. *Telematika*, 15(2), 81–91. <https://doi.org/10.35671/telematika.v15i2.1855>
- Baowaly, M. K., Sarkar, B. C., Walid, M. A. A., Ahamad, M. M., Singh, B. C., Alvarado, E. S., Ashraf, I., & Samad, M. A. (2024). Deep transfer learning-based bird species classification using mel spectrogram images. *PLoS ONE*, 19(8 August). <https://doi.org/10.1371/journal.pone.0305708>
- Chang, Y.-T., & Sinnott, R. (2022). Machine Learning-based Classification of Birds through Birdsong. *ArXiv*, abs/2212.04684. <https://doi.org/10.48550/arXiv.2212.04684>
- Elsayed, N., Maida, A. S., & Bayoumi, M. (2019). *Deep Gated Recurrent and Convolutional Network Hybrid Model for Univariate Time Series Classification*. <https://doi.org/10.14569/IJACSA.2019.0100582>
- Ferdiana, R., Dicka, W. F., & Boediman, A. (2021). Cat sounds classification with convolutional neural network. *International Journal on Electrical Engineering and Informatics*, 13(3), 755–765.
- Foggia, P., Greco, A., Roberto, A., Saggese, A., & Vento, M. (2023). Degramnet: effective audio analysis based on a fully learnable time–frequency representation. *Neural Computing and Applications*, 35(27), 20207–20219. <https://doi.org/10.1007/s00521-023-08849-7>
- Hernández-Navarro, A. J., Robledano, F., Jiménez-Franco, M. V., Royle, J. A., & Calvo, J. F. (2024). Long-term trends of local bird populations based on monitoring schemes: are they suitable for justifying management measures? *Journal of Ornithology*, 165(2), 355–367. <https://doi.org/10.1007/s10336-023-02114-3>

- Levy, J., Naitsat, A., & Zeevi, Y. Y. (2022). Classification of audio signals using spectrogram surfaces and extrinsic distortion measures. *Eurasip Journal on Advances in Signal Processing*, 2022(1). <https://doi.org/10.1186/s13634-022-00933-9>
- Myagila, K., Nyambo, D. G., & Dida, M. A. (2025). Efficient spatio-temporal modeling for sign language recognition using CNN and RNN architectures. *Frontiers in Artificial Intelligence*, 8. <https://doi.org/10.3389/frai.2025.1630743>
- Ozcan, T., & Gungor, H. (2025). Baby Cry Classification Using Structure-Tuned Artificial Neural Networks with Data Augmentation and MFCC Features. *Applied Sciences*, 15(5), 2648.
- Revadekar, S., Panchal, V., Kanani, P., Shah, K., Vasoya, A., & Pandey, R. (2023). Bird Sound Classification using Deep Neural Networks: A Comparative Analysis of State-of-the-Art Models and Custom Architectures. *International Journal of Intelligent Systems and Applications in Engineering*, 11(4), 614–622. <https://ijisae.org/index.php/IJISAE/article/view/3596>
- Zarzycki, K., & Ławryńczuk, M. (2021). LSTM and GRU neural networks as models of dynamical processes used in predictive control: A comparison of models developed for two chemical reactors. *Sensors*, 21(16). <https://doi.org/10.3390/s21165625>