

Original Article

e-ISSN: 2581-0545 - <https://journal.itera.ac.id/index.php/jsat/>

Implementasi Transfer Learning pada Deep Neural Network Untuk Identifikasi Suara Burung di Pulau Sumatera

Received 00th January 20xx
Accepted 00th February 20xx
Published 00th March 20xx

Open Access

DOI: 10.35472/x0xx0000

Ardika Satria ^{*a}, Author 2^b, Author 3^c^a Institut Teknologi Bandung^b Affiliation of Author 2^c Affiliation of Author 3* Corresponding E-mail: journal@itera.ac.id

Abstract: The conservation forest area that protects endangered species in Sumatra is only 5 million ha, while the remaining forest area in Sumatra is only 22.8 million ha. This threatens the sustainability of biodiversity in Sumatra and triggers global climate change. The impact of deforestation has made it difficult for many animals to communicate with each other. Communication is built usually through the ability to transmit acoustic signals that can be understood by fellow species. The activity of natural sounds originating from animal sounds (biophony) can be used as an indicator of changes in their habitat environment. This research was conducted by extracting acoustic data of bird sounds on the island of Sumatra with feature extraction Mels Spectrogram. Comparative analysis of various transfer learning models, ResNet, VGG-16, EfficientNet, DenseNet and Xception with two different optimizers, Adam and RMSProp. It was found that the Transfer Learning model with DenseNet got a better accuracy of 0.91 with the RMSProp optimizer on the Mels Spectrogram extraction feature.

Keywords: CNN, Bioacoustic, Transfer Learning, Sumatera

Abstrak: Kawasan hutan konservasi yang melindungi spesies langka di Sumatera hanya seluas 5 juta ha, sedangkan luas hutan di Sumatera yang tersisa hanya 22,8 juta ha. Hal ini mengancam keberlangsungan kelestarian keanekaragaman hayati di Sumatera dan memicu perubahan iklim global. Dampak deforestasi telah menyulitkan banyak satwa saling berkomunikasi satu sama lain. Komunikasi yang dibangun biasanya melalui kemampuan memancarkan sinyal akustik yang dapat dipahami sesama spesies. Aktivitas suara alam yang bersumber dari suara hewan (biofoni) tersebut dapat digunakan sebagai indikator perubahan lingkungan habitatnya. Penelitian ini dilakukan dengan mengekstrak data akustik suara burung di Pulau Sumatera dengan fitur ekstraksi Mels Spectrogram. Analisis komparasi terhadap berbagai model transfer learning yaitu, ResNet, VGG-16, EfficientNet, DenseNet dan Xception dengan dua optimizer berbeda yaitu Adam dan RMSProp. Diperoleh bahwa model Transfer Learning dengan DenseNet mendapat akurasi lebih baik mencapai 0,91 dengan optimizer RMSProp pada fitur ekstraksi Mels Spectrogram.

Kata Kunci : CNN, Bioakustik, Transfer Learning, Sumatera

Pendahuluan

Salah satu sumber daya alam yang melimpah di Indonesia adalah hutan. Kurang lebih 40 persen produksi oksigen dunia berasal dari hutan-hutan Indonesia [1]. Sementara itu, luas hutan di wilayah Indonesia saat ini telah tersisa 50,9 persen atau sekitar 95,6 juta ha [2]. Kondisi ini semakin turun dalam kurun waktu 10 tahun sebelumnya, yang disebabkan oleh deforestasi karena tumbuhnya industrialisasi hingga kebijakan yang mengaturnya. Hal tersebut menjadi ancaman besar bagi keberlangsungan kelestarian dan keanekaragaman hayati di Indonesia dan terhadap perubahan iklim di

dunia [3]. Peningkatan deforestasi dan degradasi hutan di pulau Sumatera telah menyumbang 58 persen secara nasional dalam rentang tahun 2018 hingga 2019 [2]. Meskipun tutupan hutan alam yang tersisa di pulau Sumatera saat ini adalah 22,8 juta ha [2], namun perubahan drastis tersebut sudah berada diambang yang mengkhawatirkan. Sedangkan hutan konservasi yang melindungi spesies langka di pulau Sumatera hanya seluas 5.227.282 ha yang terbagi di 14 wilayah hutan konservasi. Biofoni merupakan salah satunya suara alam yang dihasilkan oleh hewan seperti burung, mamalia, amfibi dan serangga, serta spesies lainnya, berperan penting sebagai alarm alami yang sangat sensitif



terhadap perubahan ekologi [4]. Komunikasi setiap spesies melalui sinyal akustik yang dihasilkan oleh mereka memiliki karakteristik yang berbeda-beda sehingga pemantauan melalui pengenalan suara dapat menandakan keberadaan spesies yang masih bertahan. Sedangkan pemantauan spesies di hutan konservasi masih berfokus pada menghitung jumlah keberadaan spesies.

Padahal pemantauan terhadap suara dapat membantu dalam mengidentifikasi persebaran spesies dan juga mengetahui kesehatan lingkungan habitatnya [5]. Kebanyakan spesies telah membangun relung dalam ruang lingkup akustik (frekuensi dan waktu) selama beberapa dekade akibat dari seleksi alam [6]. Relung yang terbentuk tersebut apakah hasil seleksi secara konvergen dengan mengefisiensi sinyal di lingkungan yang berisik ataupun lingkungan yang telah berubah akibat kerusakan habitat asli, atau seleksi secara divergen ke zona nyaman di relung yang berbeda, yang masih harus dikaji lebih lanjut [7]. Maka dari itu, pemahaman tentang lingkungan akustik dan *ecological stressor* menjadi sangat penting dalam pemantauan hutan secara berkala [8].

Proses pengolahan data akustik menggunakan *deep learning* yang merupakan bagian dari *machine learning*, telah memberikan banyak kontribusi pada beberapa tahun ini [9]. Sehingga mengalami banyak peningkatan pesat terutama penerapannya untuk sistem identifikasi otomatis pada bioakustik [10]. Salah satu arsitektur *deep learning* yang paling populer digunakan adalah Convolution Neural Network (CNN) [9]. Aplikasi dari CNN banyak diterapkan pada pengolahan citra dan pengenalan pola. Struktur jaringan CNN memiliki karakteristik layer yang didesain khusus agar dapat memproses varian *feature extraction* dari objek citra [11]. Sedangkan untuk pengenalan suara, input yang digunakan pada model CNN berupa visualisasi suara seperti Mel-Frequencies Cepstral Coefficient (MFCC) ataupun model spektrogram statistik lainnya [12]. Berdasarkan kontes BirdCLEF, model CNN dengan *single layer* telah berhasil mengklasifikasi banyak spesies burung berdasarkan citra gambar dan data akustik pemantauan hutan konservasi [13]. Model tersebut merupakan improvisasi dari model sebelumnya yang masih mengandalkan model berbasis Support Vector Machine [14], Decision Tree dan Random Forest [15]. Yang mana model sebelumnya tidak lagi mampu menggunakan *feature extraction* yang lebih besar sehingga dilakukan pengurangan data agar menghemat waktu dan beban komputasi [10].

Sprengel, dkk. menggunakan enam layer dan *single-label* CNN pada rekaman suara burung untuk mengklasifikasi berbagai spesies burung pada kontes BirdCLEF 2016 [16]. Ia juga menerapkan teknik augmentasi yaitu kombinasi sampel audio pada kelas yang sama dengan sampel noise agar performa CNN dapat meningkat [16]. Pada penelitian Ruff, dkk. menunjukkan bahwa dengan enam layer pada CNN dapat mengidentifikasi dan mengklasifikasi enam suara burung hantu dengan tambahan data noise yang tidak mengandung *class* [17]. J.LeBien, dkk. memodifikasi model CNN agar dapat memproses multi-label pada rekaman suara burung dan katak, sehingga menghasilkan akurasi yang lebih tinggi [18]. Sedangkan hasil penelitian Z.J. Ruff dkk., telah berhasil membangun *workflow* sistem otomatis dengan enam layer pada CNN juga berhasil mengklasifikasi berbagai suara hewan antara lain burung, amfibi dan mamalia [9]. Penelitian Hidayat dkk., menunjukkan bahwa dengan menggabungkan dua input fitur ekstraksi Mels Spectrogram dan MFCC pada CNN menghasilkan mean average precision hingga 97 persen [19]. Xie dk., melakukan analisis suara katak dengan model sederhana CNN yang menggunakan fitur ekstraksi multi-view spectrogram dengan fungsi twinloss menghasilkan akurasi hingga 99 persen [20].

Pada penelitian ini dilakukan klasifikasi suara yang di hasilkan dari berbagai spesies burung yang berasal dari Pulau Sumatera dengan menggunakan deep learning melalui penerapan transfer learning. Adapun transfer learning yang digunakan adalah ResNet 50, VGG-16, EfficientNet, DenseNet dan Xception. Analisis komparasi model transfer learning dilakukan dengan optimasi yang berbeda, yaitu Adam, RMSProp, dan SGD untuk mendapatkan hasil akurasi terbaik dan model yang presisi.

Metode

Dataset

Pengambilan data dilakukan dengan menggunakan query yang telah disediakan oleh API *default* dari platform xenocanto. Query tersebut memuat nama spesies dari region pulau Sumatera, jenis suara seperti *call* dan *song*, dan kualitas audio. Kemudian dipilih juga audio dari berbagai region dari luar pulau Sumatera dengan identifikasi spesies yang sama. Terdapat 13 spesies burung yang memiliki kelimpahan data diatas 50 buah dengan rata-rata kualitas berada di skor paling baik seperti yang ditunjukkan oleh Tabel 1. Adapun audio yang telah didapat memiliki banyak variasi panjang durasi, mulai dari 5 detik hingga beberapa menit. Lalu

dilakukan pemotongan durasi sepanjang 5 detik agar memiliki panjang durasi yang sama. Data audio yang sudah sama akan diproses dengan fitur ekstraksi mels spectrogram. `Selanjutnya data dibagi menjadi 80% data pelatihan, 10% data validasi dan 10% data test.

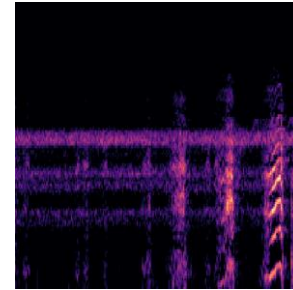
Tabel 1. Jumlah data audio spesies burung

No	Nama Umum	Nama Spesies	Jumlah file
1	White-bellied Woodpecker	<i>Dryocopus javensis</i>	128
2	Large-tailed Nightjar	<i>Caprimulgus macrurus</i>	190
3	Pygmy Cupwing	<i>Pnoepyga pusilla</i>	214
4	Rufous-tailed Tailorbird	<i>Orthotomus sericeus</i>	122
5	Horsfield's Babbler	<i>Malacocincla sepiaria</i>	108
6	Sooty-capped Babbler	<i>Malacopteron affine</i>	86
7	Mangrove Whistler	<i>Pachycephala cinerea</i>	73
8	Slender-billed Crow	<i>Corvus enca</i>	122
9	Arctic Warbler	<i>Phylloscopus borealis</i>	371
10	Pin-striped Tit-Babbler	<i>Macronus gularis</i>	252
11	Rufous-browed Flycatcher	<i>Anthipes solitaris</i>	57
12	Black-winged Kite	<i>Elanus caeruleus</i>	79
13	Oriental Pied Hornbill	<i>Anthracoceros albirostris</i>	256
Total			2063

Data Preprocessing

Pemrosesan audio pada fitur ekstraksi mels spectrogram, yang mana dilakukan transformasi data dari data berupa audio berformat mp3 menjadi klasifikasi gambar. Mels Spectrogram diperoleh dari STFT (short time fourier transform) dengan *overlapping window segment*. Pada setiap segmen ukuran windownya adalah 512, dan ukuran hop size adalah 128. Pada proses ini dilakukan konversi dari domain waktu menjadi domain frekuensi. Kemudian DTFT (Discrete Time Fourier Transform) digunakan untuk mentransformasi dari log scale menjadi mels spectrogram yang selanjutnya digunakan 128 Mels untuk hasil mels spectrogram. Fungsi window yang digunakan adalah Hann window dengan NFFT sebesar 1024. Parameter lain yang digunakan adalah sample rate yaitu 48 kHz. audio di konversi dalam decibel (dB) di ubah dimensinya menjadi 224 x 224 *RGB channels* dan

disimpan sebagai gambar yang berwarna. Salah satu mels spectrogram dari spesies *Corvus enca* di tunjukkan pada **Gambar 1**.



Gambar 1. Contoh Mels Spectrogram

Selanjutnya tahap data processing dengan fitur mels spectrogram, dengan memotong audio menjadi 5 detik. Parameter yang digunakan untuk menghasilkan melspectrogram adalah, $n_fft = 1024$, $hop_size = 1024$, $n_mels = 128$, $win_size = 512$, window type = hann, $fmin = 1400$. Dengan $n_mels = 128$, yang merupakan high pass filter untuk mengurangi noise pada audio. Adapun framework yang diaplikasikan pada data preprocessing adalah Librosa. Untuk framework model, digunakan TensorFlow Keras dan visualisasi dengan matplotlib.

Model 1 – VGG16

Arsitektur model VGG16 untuk mengklasifikasi 13 kelas spesies burung digunakan gambar RGB sebagai input dengan ukuran 224 x 224 x 3. Kemudian melewati 5 layer convolutional dan tiga *fully-connected layer*. Fungsi aktivasi yang digunakan pada setiap layer adalah rectified linear unit (ReLU) dan Max pooling layer diterapkan pada setiap akhir layer. Dua *fully-connected layer* terakhir dengan 4096 ReLU dan layar terakhir dengan fungsi aktivasi *softmax*. Pada saat proses training diterapkan optimasi Adam dan RMSProp.

Model 2 – ResNet50

Model ini pertama kali dikembangkan oleh He dkk. yang mana ia menggunakan 'koneksi jalan pintas' untuk meningkatkan *convergence rate* dan akurasi pada klasifikasi [21]. ResNet50 digunakan sebagai pre-trained transfer learning dengan weight yang dipakai adalah imagenet. Kemudian ditambahkan fully connected layer sebagai fine-tuned parameter dari arsitektur ResNet50. Terdapat 4 fully connected layer dengan fungsi aktivasi ReLU, dan layar output dengan fungsi aktivasi softmax.

Model 3 – EfficientNet

Model ini merupakan salah satu model CNN yang menggunakan metode skala yang seragam untuk semua dimensi dari resolusi gambar dengan compound coefficient (ϕ). tiga dimensi seperti, depth, width dan resolution yang dirumuskan sebagai berikut: $depth = d = \alpha^\phi$, $width = w = \beta^\phi$, $resolution = r = \gamma^\phi$. Di mana nilai coefficient compound ditentukan oleh user. Model ini dapat mengimbangi perubahan pada ukuran proposional pada model CNN yang mana jika parameter depth di besarkan maka parameter lain akan dua kali lebih besar, untuk meminimalkan terjadinya ledakan gradien, maka digunakan coefficient compound. Salah satu arsitektur EfficientNet yang digunakan adalah EfficientNetB4. Weight yang digunakan adalah imagenet.

Model 4 – DenseNet

Arsitektur DenseNet memiliki karakteristik yang disebut dengan dense block yang mana pada setiap blok di setiap layer terhubung secara langsung. Sebuah layer mengambil input dari output keseluruhan layer sebelumnya yang memberikan output ke semua layer setelahnya sehingga jaringan menjadi lebih ramping. Salah satu arsitektur sederhana dari DenseNet adalah DenseNet169. Berdasarkan pustaka keras DenseNet169 memiliki 427 layer yang terbagi dalam 58 dense blok. Kemudian dihubungkan dengan empat fully connected layer dengan fungsi aktivasi ReLU, dan fungsi aktivasi outputnya adalah softmax.

Model 5 – Xception

Model Xception merupakan model extrem dari Inception. Dimana lebih efisien dengan mengaplikasikan filter setiap kedalaman jaringan kemudian mengompresi dengan menggunakan konvolusi input 1x1 pada keseluruhan jaringan. Weight yang digunakan adalah imagenet yang dihubungkan dengan empat fully connected layer dan fungsi aktivasi ReLU. Sedangkan fungsi output yang digunakan masih sama yaitu softmax.

Hyperparameter

Parameter dijaga stabil dengan dropout layer sebesar 50% dengan 10 epoch, dan batch yang digunakan adalah 16. Learning rate dijaga 0,0001. sedangkan parameter Adam dengan decay senilai 10^{-7} , dan RMSprop dengan momentum sebesar 0,9. Pada model ResNet 50 dan DenseNet169 dilakukan variasi yang lebih banyak dengan dropout layer 20% dan learning rate sebesar 0,0001 dengan Batch yang dipakai sebanyak 16. Kemudian di uji kembali dengan dropout tetap dan learning rate sebesar 0,00001 serta Batch yang digunakan sebanyak 32. Uji selanjutnya pada model ini adalah digunakan parameter yang sama seperti sebelumnya

dengan dropout sebesar 50%. Keempat parameter yang berbeda tersebut diuji dengan optimisasi adam dan rmsprop. Sehingga jumlah model yang digunakan dalam penelitian ini sebanyak 22 model. Seperti yang ditunjukkan pada Tabel 2.

Tabel 2. Banyaknya Model dan Hyperparameter yang digunakan.

No	Model	Dropout	Learning Rate	Batch
1	VGG16	20%	0,0001	16
2	EfficientNet	20%	0,0001	16
3	Xception	20%	0,0001	16
4	ResNet50	20%	0,0001	16
5	DenseNet169	20%	0,00001	32
		50%	0,0001	16
		50%	0,00001	32
		20%	0,0001	16
		20%	0,00001	32
		50%	0,0001	16
		50%	0,00001	32
		50%	0,00001	32

Metrik Evaluasi

Yang mana R dan P merupakan Recall dan Precision, yaitu [9]:

$$P(\tau) = \frac{TP}{N'_p} = \frac{TP}{TP + FP} \quad (1)$$

$$R(\tau) = \frac{TP}{N_p} = \frac{TP}{TP + FN} \quad (2)$$

Di mana $TP = True\ Positive$ yang merepresentasikan prediksi yang sesuai dengan validasi data, $TN = True\ Negative$ yang menunjukkan prediksi benar tidak terdapat suara hewan, $FP = False\ Positive$ yang merupakan bentuk kesalahan prediksi pada suara hewan yang tidak ada tapi dikategorikan ada, dan $FN = False\ Negative$, yang merupakan bentuk kesalahan prediksi suara hewan yang ada tetapi di prediksi tidak ada. Sedangkan N'_p dan N_p adalah jumlah data prediksi dan jumlah data target. Recall atau disebut juga sensitivitas digunakan untuk melihat seberapa besar laju *true-positive* yang dihasilkan. Selain itu pengukuran evaluasi dapat dilakukan melalui analisis AUCROC atau Area Under Curve the Receiver Operating Characteric [22]. AUCROC digunakan untuk mengukur ekspektasi antara data yang diprediksi benar (TP) yang tersebar secara merata dari pada hasil prediksi FP ataupun FN. Kurva ROC adalah representasi dalam bentuk grafik dari hubungan antara sensitivitas dan presisi. Kurva ROC banyak digunakan untuk menggambarkan keakuratan sebuah model klasifikasi [23]. Nilai AUCROC yang diharapkan adalah 1,0 yang tersebar merata dan konsisten sepanjang grafik .

Hasil dan Pembahasan

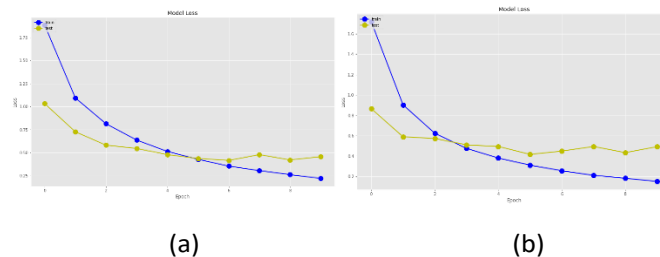
Pada Penelitian ini menggunakan perbandingan data 80:10:10 untuk data pelatihan (training), data validasi (validation), dan data test (test). Ditemukan sebanyak 10.590 data gambar pada pelatihan dan 1.303 data pada data validasi dan test. Komputasi pemrosesan citra dibantu oleh GPU dari google colab pro sebanyak 1 GPU. Rata-rata waktu komputasi menghabiskan 8 jam untuk sekali mengeksekusi program pada satu model.

Perbandingan akurasi dan loss data test pada kelima model ditunjukkan pada Tabel 3. Dimana pada model ini diterapkan parameter pertama, yaitu dropout 50%, learning rate sebesar 0,0001 dan batch size sebesar 16.

Tabel 3. Perbandingan Akurasi dan Loss

No	Model	Adam		RMSProp	
		Akurasi	Loss	Akurasi	Loss
1	EfficientNet B0	0,56	1,56	0,80	0,79
2	Densenet169	0,89	0,47	0,91	0,39
3	Xception	0,88	0,48	0,89	0,49
4	ResNet50	0,88	0,50	0,88	0,59
5	Vgg16	0,84	0,77	0,86	0,49
6	EfficientNet B4	0,83	0,60	0,84	0,60

Hasil terbaik ditunjukkan pada model Densenet169, dengan nilai akurasi mencapai 0,91 atau 91% dengan nilai loss yang kecil yaitu 0,39. Pada kedua optimasi yang digunakan menunjukkan bahwa optimisasi RMSProp lebih cepat dan tidak mengalami ledakan gradien, yang ditunjukkan dengan loss yang kecil dan nilai akurasi yang lebih baik dibandingkan dengan optimisasi adam. Urutan kedua terbaik adalah pada model Xception dimana nilai akurasi yang didapat mencapai 89% dengan nilai loss 0,49. Dibandingkan dengan model sebelumnya yaitu densenet169, model xception memiliki keunggulan dimana pada epoch ke 5 bahwa akurasi pelatihan dan test tidak mengalami lonjakan gradien yang membesar dan hanya terjadi overfitting yang kecil dan stabil. Perbandingan model loss optimisasi adam dan RMSprop ditunjukkan pada Gambar 2 dan 3.



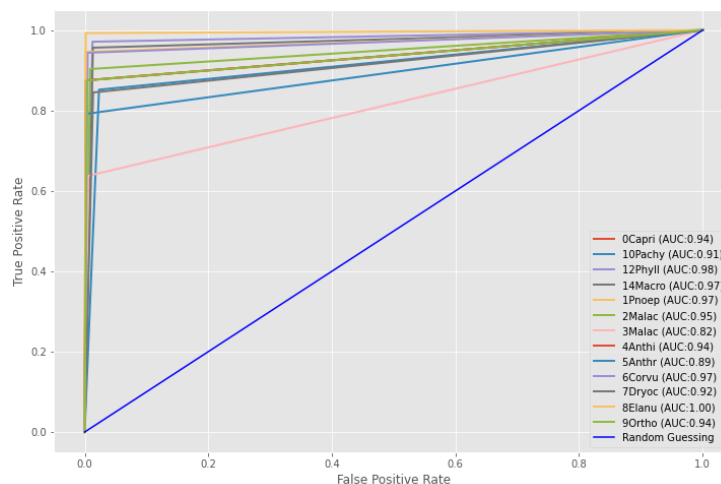
Gambar 3. (a) Adam dan (b) RMSProp model loss Xception

Kemudian perbandingan nilai F1 score dan AUCROC ditunjukkan pada Tabel 4

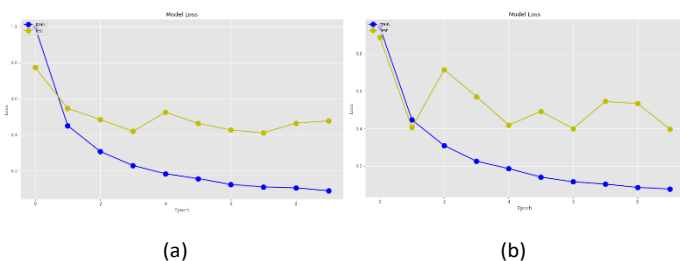
Tabel 4. Perbandingan Nilai F1 Score dan AUCROC

No	Model	Adam		RMSProp	
		F1 Score	AUCROC	F1 Score	AUCROC
1	EfficientNet B0	0,57	0,70	0,80	0,84
2	Densenet169	0,90	0,92	0,91	0,93
3	Xception	0,89	0,91	0,90	0,92
4	ResNet50	0,89	0,91	0,89	0,91
5	Vgg16	0,85	0,89	0,87	0,90
6	EfficientNet B4	0,83	0,85	0,84	0,89

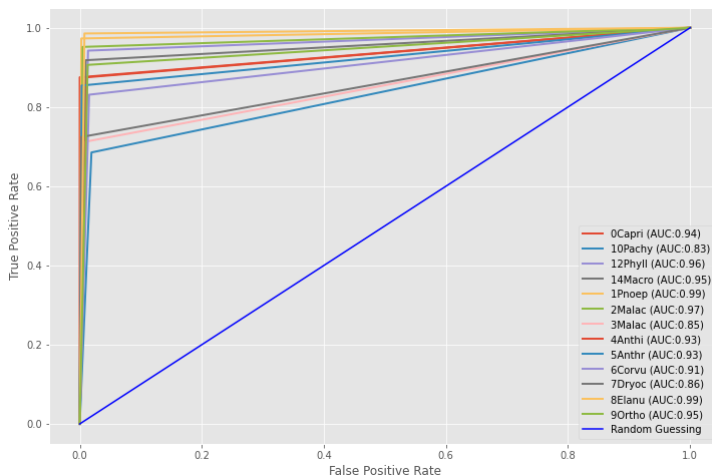
Hasil terbaik didapat oleh model Densenet169 dengan F1 score sebesar 0,91 dan AUCROC sebesar 0,93. Pada urutan kedua diperoleh model Xception dengan F1 score sebesar 0,90 dan AUCROC sebesar 0,92. Nilai F1 score yang tinggi menunjukkan bahwa nilai presisi dan recall berada diatas 0,70. Sedangkan pada F1 score model Xception masih terdapat nilai dibawah 0,70 pada beberapa kelas. Kurva AUCROC pada kedua model terbaik ditunjukkan pada Gambar 4.



Gambar 4. Kurva AUCROC dengan Optimisasi RMSProp pada Model Densenet169



Gambar 2. (a) Adam dan (b) RMSProp model loss Densenet169



Gambar 5. Kurva AUCROC dengan Optimisasi RMSProp pada Model Xception

Kemudian pada parameter kedua, hasil terbaik didapat optimisasi adam pada model densenet dengan dropout 50%, learning rate sebesar 0,0001 dan batch size sebesar 16. Seperti yang ditunjukkan pada Tabel 5.

Tabel 5. Hasil Dengan Optimisasi Adam Pada Model Resnet50 dan Densenet169

Model	Dropout	Learning rate/Batch size	Akurasi	Loss	AUCROC	F1 Score
ResNet50	0,2	0,0001/16	0.8810	0.4456	0.9163	0.88
	0,2	0,00001/32	0.8749	0.4681	0.9127	0.87
	0,5	0,0001/16	0.8711	0.5098	0.9006	0.87
	0,5	0,00001/32	0.8780	0.4378	0.9116	0.88
Densenet 169	0,2	0,0001/16	0.8915	0.3951	0.9234	0.89
	0,2	0,00001/32	0.8665	0.4719	0.9076	0.87
	0,5	0,0001/16	0.9071	0.3671	0.9347	0.91
	0,5	0,00001/32	0.8772	0.9331	0.9118	0.88

Lalu hasil terbaik optimisasi RMSProp didapat pada model densenet dengan dropout 20%, learning rate sebesar 0,0001 dan batch size sebesar 16. Seperti yang ditunjukkan pada Tabel 6.

Perbedaan yang terdapat parameter struktur jaringan pada model densenet, dengan dropout sebesar 50% pada optimisasi adam ditemukan loss yang lebih kecil dibandingkan pada optimisasi RMSprop dengan dropout sebesar 20%. Nilai AUCROC ditemukan sama dengan kedua model. Sama halnya dengan nilai F1 score yang tetap stabil pada nilai 0,91 atau 91%. Sedangkan akurasi pada kedua model mencapai nilai 90%, artinya presisi dan recall yang didapat juga seimbang dan stabil.

Tabel 6. Hasil Dengan Optimisasi RMSProp Pada Model Resnet50 dan Densenet169

Model	Dropout	Learning rate/Batch size	Akurasi	Loss	AUCROC	F1 Score
ResNet50	0,2	0,0001/16	0.8565	0.5854	0.8934	0.86
	0,2	0,00001/32	0.8626	0.5171	0.8979	0.86
	0,5	0,0001/16	0.8987	0.388	0.9324	0.9
	0,5	0,00001/32	0.8703	0.516	0.9103	0.87
Densenet 169	0,2	0,0001/16	0.9064	0.4603	0.9357	0.91
	0,2	0,00001/32	0.8956	0.3742	0.9252	0.9
	0,5	0,0001/16	0.8987	0.4499	0.9328	0.9
	0,5	0,00001/32	0.8864	0.4279	0.9191	0.89

Kesimpulan

Adapun kesimpulan dari penelitian ini adalah Model DenseNet169 merupakan model terbaik untuk klasifikasi suara burung di Pulau Sumatera, karena lebih cepat dengan 10 epoch dan akurasi yang lebih tinggi yaitu 91%. Optimisasi RMSprop lebih cepat konvergen menuju nol di dibandingkan dengan Optimisasi Adam untuk klasifikasi suara burung pulau Sumatera. Data Augmentasi berhasil menaikkan akurasi untuk fitur Melspectrogram. Klasifikasi Multiclass data suara dengan fitur melspectrogram dapat menggunakan optimisasi RMSprop untuk meningkatkan akurasi. AUC ROC score pada optimisasi RMSprop dapat menurunkan false positive rate. Penelitian selanjutnya dapat melakukan klasifikasi dengan variasi fitur ekstraksi seperti MFCC, chirplet spektrogram.

Ucapan Terimakasih

Ucapan terima kasih kepada Bapak Acep Purqon, Ph.d, dan Bapak Dr. Attar Luqman, yang telah memberikan arahan dan bimbingan hingga selesainya penulisan paper ini.

References

- [1] Michael. G, "Tropical Rainforest." 2002.
- [2] Subdirektorat Statistik Kehutanan, "Statistik Produksi Kehutanan Indonesia 2019," no. Katalog Nomor 5601005, Jakarta: Badan Pusat Statistik, 2019.
- [3] T. Garg, "Ecosystems and human health: The local benefits of forest cover in Indonesia," *J. Environ. Econ. Manage.*, vol. 98, p. 102271, 2019.

- [4] A. Farina, *Soundscape ecology: Principles, patterns, methods and applications*. Urbino: Springer, 2014.
- [5] G. Easteria, "Biofoni di Hutan Akasia, Taman Nasional Gunung Merapi, Yogyakarta." Universitas Gadjah Mada, 2017.
- [6] B. L. Krause, "The niche hypothesis: a virtual symphony of animal sounds, the origins of musical expression and the health of habitats," *Soundscape Newsl.*, vol. 6, pp. 6–10, 1993.
- [7] J. A. Tobias, R. Planqué, D. L. Cram, and N. Seddon, "Species interactions and the structure of complex communication networks," *Proc. Natl. Acad. Sci.*, vol. 111, no. 3, pp. 1020–1025, 2014.
- [8] M. V Lomolino, B. C. Pijanowski, and A. Gasc, "The silence of biogeography," *J. Biogeogr.*, vol. 42, no. 7, pp. 1187–1196, 2015.
- [9] Z. J. Ruff, D. B. Lesmeister, C. L. Appel, and C. M. Sullivan, "Workflow and convolutional neural network for automated identification of animal sounds," *Ecol. Indic.*, vol. 124, p. 107419, 2021.
- [10] 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.
- [11] C.-C. Kao, W. Wang, M. Sun, and C. Wang, "R-CRNN: Region-based convolutional recurrent neural network for audio event detection," *arXiv Prepr. arXiv1808.06627*, 2018.
- [12] N. Priyadarshani, S. Marsland, and I. Castro, "Automated birdsong recognition in complex acoustic environments: a review," *J. Avian Biol.*, vol. 49, no. 5, p. jav-01447, 2018.
- [13] A. Gomez Villa, A. Salazar, and F. Vargas, "Towards automatic wild animal monitoring: Identification of animal species in camera-trap images using very deep convolutional neural networks," *Ecol. Inform.*, vol. 41, pp. 24–32, 2017, doi: <https://doi.org/10.1016/j.ecoinf.2017.07.004>.
- [14] A. Joly, V. Leveau, J. Champ, and O. Buisson, "Shared nearest neighbors match kernel for bird songs identification-lifeclef 2015 challenge," in *CLEF: Conference and Labs of the Evaluation Forum*, 2015, no. 1391.
- [15] D. Stowell and M. D. Plumbley, "Audio-only bird classification using unsupervised feature learning," 2014.
- [16] E. Sprengel, M. Jaggi, Y. Kilcher, and T. Hofmann, "Audio based bird species identification using deep learning techniques," 2016.
- [17] Z. J. Ruff, D. B. Lesmeister, L. S. Duchac, B. K. Padmaraju, and C. M. Sullivan, "Automated identification of avian vocalizations with deep convolutional neural networks," *Remote Sens. Ecol. Conserv.*, vol. 6, no. 1, pp. 79–92, 2020.
- [18] J. LeBien et al., "A pipeline for identification of bird and frog species in tropical soundscape recordings using a convolutional neural network," *Ecol. Inform.*, vol. 59, p. 101113, 2020.
- [19] A. A. Hidayat, T. W. Cenggoro, and B. Pardamean, "Convolutional Neural Networks for Scops Owl Sound Classification," *Procedia Comput. Sci.*, vol. 179, pp. 81–87, 2021, doi: <https://doi.org/10.1016/j.procs.2020.12.010>.
- [20] J. Xie, M. Zhu, K. Hu, J. Zhang, H. Hines, and Y. Guo, "Frog calling activity detection using lightweight CNN with multi-view spectrogram: A case study on Kroombit tinker frog," *Mach. Learn. with Appl.*, vol. 7, p. 100202, 2022, doi: <https://doi.org/10.1016/j.mlwa.2021.100202>.
- [21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [22] V. Bewick, L. Cheek, and J. Ball, "Statistics review 13: receiver operating characteristic curves," *Crit. care*, vol. 8, no. 6, pp. 1–5, 2004.
- [23] S. Ekelund, "Roc Curves—What are they and how are they used?," *Point Care*, vol. 11, no. 1, pp. 16–21, 2012.