



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

Open Access

DOI: 10.35472/x0xx0000

Klasifikasi Suara Katak Menggunakan Model Deep Learning Modified DenseNet-121 dan DenseNet-169 dengan Fitur Ekstraksi MFCC

Kharisma Gumilang 1 ^{*a}, Afwa Fuadi Nugraha 2 ^b,
Isyamaetreyra Savitri 3 ^c, Nadia Fitri Yani 4 ^d, Yunita
Amelia Puspitasari 5 ^e, Jasmine Georgina Sekartaji
6 ^f

^{a, b, c, d, e, f}Program Studi Sains Data, Institut Teknologi Sumatera

* Corresponding E-mail: kharisma.121450142@student.itera.ac.id

Abstract: Frogs, as one of the amphibian species, have many distinctive characteristics that make them fascinating creatures to study. The diversity of frog species, the unique sounds they produce, and their important role in the ecosystem make them effective bioindicators for environmental monitoring and biodiversity conservation. Each frog species produces a different sound, which can be used for identification without direct observation. This study aims to classify frog sounds using deep learning technology, with Mel-frequency Cepstral Coefficients (MFCC) features that have been proven effective in sound representation. The dataset used includes 1512 sound recordings from 9 frog species: *Dendrosophus minutus*, *Leptodactylus fuscus*, *Rana temporaria*, *Rhinella marina*, *Scinax rubex*, Boana, Pepper frog, Pool frog, and South frog. The data was divided into 80% for training, 10% for validation, and 10% for testing. Two deep learning models were used, namely Modified DenseNet-121 and Modified DenseNet-169. The models were trained using the Adam algorithm with a learning rate of $1e-4$, weight decay of $1e-4$, and the ReduceLROnPlateau mechanism for learning rate adjustment. The results of Modified DenseNet-121 showed an accuracy of 68%, precision of 71%, recall of 68%, and F1 score of 68%. Meanwhile, Modified DenseNet-169 resulted in an accuracy of 61%, precision of 68%, recall of 61%, and F1 score of 62%. This approach is expected to assist in biodiversity monitoring efforts and species conservation by utilizing deep learning technology.

Keywords: Deep Learning, DenseNet, MFCC, Sound Classification, Conservation

Abstrak: Katak merupakan salah satu spesies amfibi, memiliki banyak ciri khas yang menjadikannya makhluk yang menarik untuk dipelajari. Keanekaragaman spesies katak, suara unik yang dihasilkan, serta peran pentingnya dalam ekosistem menjadikannya sebagai bioindikator yang efektif dalam pemantauan lingkungan dan konservasi keanekaragaman hayati. Setiap spesies katak menghasilkan suara yang berbeda, yang dapat digunakan untuk identifikasi tanpa observasi langsung. Penelitian ini bertujuan untuk mengklasifikasikan suara katak menggunakan teknologi *deep learning*, dengan fitur *Mel-frequency Cepstral Coefficients* (MFCC) yang telah terbukti efektif dalam representasi suara. Dataset yang digunakan



mencakup 1512 rekaman suara dari 9 spesies katak, yaitu *Dendrosophus minutus*, *Leptodactylus fuscus*, *Rana temporaria*, *Rhinella marina*, *Scinax rubex*, *Boana*, *Pepper frog*, *Pool frog*, dan *South frog*. Data dibagi menjadi 80% untuk pelatihan, 10% untuk validasi, dan 10% untuk pengujian. Dua model *deep learning* yang digunakan, yaitu *Modified DenseNet-121* dan *Modified DenseNet-169*. Model dilatih menggunakan algoritma Adam dengan *learning rate* 1e-4, *weight decay* 1e-4, dan mekanisme *ReduceLROnPlateau* untuk penyesuaian *learning rate*. Hasil dari *Modified DenseNet-121* menunjukkan akurasi 68%, *precision* 71%, *recall* 68%, dan *F1 score* 68%. Sementara itu, *Modified DenseNet-169* menghasilkan akurasi 61%, *precision* 68%, *recall* 61%, dan *F1 score* 62%. Pendekatan ini diharapkan dapat membantu upaya pemantauan keanekaragaman hayati dan konservasi spesies dengan memanfaatkan teknologi *deep learning*.

Kata Kunci : *Deep Learning*, *DenseNet*, MFCC, Klasifikasi Suara, Konservasi

Pendahuluan

Katak merupakan hewan amfibi yang memiliki keanekaragaman spesies. Di Indonesia, terdapat lebih dari 10% jenis katak dari total katak di seluruh dunia atau terdapat lebih dari 7000 spesies katak di Indonesia [1]. Katak, sebagai salah satu kelompok amfibi yang memiliki peran penting dalam ekosistem. Katak berfungsi sebagai predator alami serangga dan berperan juga sebagai bagian dari rantai makanan yang mendukung keseimbangan ekosistem [2]. Keberadaan katak sering menjadi indikator utama untuk perubahan kualitas lingkungan dan habitat atau dapat dikatakan bahwa katak adalah bioindikator yang efektif dalam pemantauan ekosistem [3]. Katak adalah salah satu spesies hewan yang sangat sensitif terhadap perubahan lingkungan, misalnya polusi, deforestasi, dan perubahan iklim [4]. Dengan keberagaman biologisnya, katak sering dijadikan penelitian untuk mendalami dampak perubahan lingkungan terhadap kelangsungan hidup spesies ini.

Namun, populasi katak di berbagai wilayah dunia kini menghadapi ancaman serius akibat penangkapan lebih, hilangnya hutan dan lahan basah, penyakit, spesies introduksi dan kecacatan [5]. Pada tahun 1998, salah satu jenis katak, yaitu

katak emas di Costa Rica menjadi salah satu bukti bahwa terjadi krisis global pada populasi amfibi [5]. Pada tahun 2004, 5743 spesies amfibi, 32% spesies terancam punah, 9 spesies punah dan 43% spesies lainnya mengalami penurunan populasi [5]. Dari data ini, terdapat kekhawatiran karena amfibi salah satunya katak memiliki peran yang penting sebagai bioindikator kesehatan lingkungan, pengendalian hama serangga, sumber protein, dan bahan penelitian biomedis. Hal inilah yang menjadi penekanan penting untuk peneliti dan upaya konservasi terhadap spesies amfibi khususnya katak, supaya melindungi keberadaannya dan menjaga keseimbangan ekosistem yang lebih luas.

Untuk mendeteksi gangguan ekosistem, misalnya perubahan kualitas air, suhu, tingkat pencemaran dan kerusakan habitat dapat menggunakan suara katak. Suara katak ini dapat berguna untuk memantau kesehatan lingkungan [6]. Setiap spesies memiliki pola suara yang berbeda, sehingga suara katak ini menjadi alat yang efektif untuk identifikasi tanpa memerlukan observasi langsung. Pada saat ini, kemajuan teknologi khususnya *deep learning* dapat melakukan pengolahan data bioakustik. *Deep learning* dapat melakukan pengenalan dan klasifikasi spesies berdasarkan suara dengan tingkat akurasi yang

tinggi, bahkan di lingkungan yang terdapat gangguan.

Teknik *Mel-frequency Cepstral Coefficients* (MFCC) merupakan salah satu teknik yang digunakan pada tahap awal pengolahan suara khususnya pada *deep learning*. MFCC adalah salah satu teknik untuk mengolah data suara yang paling akurat pada fitur enkripsi dalam berbagai aplikasi secara global [7]. Dengan menggunakan MFCC, data suara akan diubah frekuensi domainnya menjadi mirip dengan cara manusia mendengar. Dengan menggunakan teknik MFCC dapat menghasilkan fitur yang relevan untuk dianalisis lebih lanjut [7]. Pada model *deep learning modified*, misalnya *DesNet-121* dan *DesNet-169* bisa menangani tantangan analisis suara yang kompleks. Modifikasi pada *DenseNet-121* dan *DenseNet-169*, seperti penyesuaian jumlah lapisan, ukuran kernel, atau integrasi mekanisme perhatian (*attention mechanism*), memungkinkan model ini mengolah fitur MFCC dengan presisi tinggi, menangkap pola kompleks dalam data suara, serta meningkatkan akurasi dan efisiensi analisis suara pada berbagai aplikasi, termasuk klasifikasi suara, identifikasi spesies, dan analisis ekosistem berbasis suara.

Dari penelitian sebelumnya yang dilakukan oleh Z.J Ruff dkk., dengan menggunakan metode CNN berhasil mengklasifikasikan 14 spesies hutan, khususnya burung dan mamalia serta berhasil membangun *workflow* sistem otomatis dengan enam layer pada CNN [8]. S. Wei dkk., menggunakan ekstraksi fitur *Log-mel spectrogram* mampu menangkap informasi waktu dan frekuensi secara lebih efektif, sehingga meningkatkan kinerja model Convolutional Neural Networks (CNN) dalam mengklasifikasikan audio dengan akurasi yang tinggi [9]. Nanni dkk., pada penelitiannya data audio yang dilatih dengan metode augmentasi data dan representasi audio,

seperti *Discrete Gabor Transform* (DGT) dan *Mel spectrogram*, mampu meningkatkan akurasi klasifikasi suara secara signifikan, yaitu akurasi hingga 26.75% [10].

Pada penelitian ini bertujuan untuk mengembangkan metode klasifikasi suara katak menggunakan teknologi *deep learning*, dengan fokus pada analisis suara dari berbagai spesies suara katak yang tersebar di habitat alamnya. Dengan menggunakan model *DesNet-121* dan *DesNet-169* yang akan dilakukan komparasi dengan optimasi Adam untuk mendapatkan hasil akurasi yang terbaik dan model yang presisi.

Metode

Dataset

Dalam penelitian ini, data katak digunakan sebagai bahan analisis untuk memahami variasi suara katak. Proses pengambilan data dilakukan dengan memanfaatkan *query* yang ada dari API default *platform* Xenocanto. Terdapat 9 spesies katak yang digunakan pada penelitian ini, seperti yang tertera pada Tabel 1. Data rekaman suara katak yang dikumpulkan memiliki variasi durasi yang berbeda-beda, mulai dari 5 detik sampai beberapa menit. Data awal yang terkumpul sebanyak 75 data suara katak yang kemudian dilakukan pemotongan durasi sepanjang 5 detik supaya memiliki panjang durasi yang sama, sehingga data suara menjadi sebanyak 1512 data suara katak. Data suara yang sudah melalui tahap preprocessing kemudian, dari format mp3 dikonversi ke format *Waveform Audio File Format* (WAV). Selanjutnya, data dibagi menjadi 80% untuk pelatihan, 10% untuk data validasi dan 10% untuk data pengujian.

Tabel 1. Dataset

No	Jenis Katak	Jumlah File
1	Boana Cinerascens	172
2	Papper Treefrog	206
3	Pool Frog	133
4	South American White-lipped Grassfrog	93
5	Dendropsophus Minutus	132
6	Rana Temporaria	426
7	Rhinella Marina	139
8	Leptodactylus Fuscus	93
9	Scinax Ruber	133
Total		1512

Data Preprocessing

Data preprocessing merupakan tahapan yang dilakukan pada data mentah sebelum digunakan dalam analisis atau pemodelan. Tujuan utama dari preprocessing adalah untuk memperbaiki kesalahan dalam data mentah, menghilangkan elemen yang tidak relevan, dan mengubah data menjadi format yang lebih mudah dipahami serta digunakan. Beberapa langkah yang dilakukan dalam preprocessing untuk data klasifikasi suara katak dalam analisis ini antara lain:

1. Segmentasi Data

Pada tahap segmentasi data, file audio dibagi menjadi segmen-segmen berdurasi 5 detik untuk

mempermudah analisis lebih lanjut. Proses dimulai dengan membuat file audio, dilanjutkan dengan pengurangan noise menggunakan metode noise reduction untuk meningkatkan kejernihan suara. Selanjutnya, volume audio diperbesar untuk meningkatkan kualitas sinyal. Durasi total audio dihitung untuk menentukan jumlah segmen yang dapat dibuat, dengan durasi masing-masing 5 detik. Audio kemudian dibagi sesuai durasi yang telah ditentukan. Segmen-segmen ini dikelompokkan berdasarkan nama kelas yang tercantum dalam nama file dan disimpan dengan penamaan yang teratur. Selain itu, jumlah file pada setiap kelas dihitung dan ditampilkan untuk memberikan gambaran mengenai distribusi data. Hasil segmentasi menunjukkan bahwa kelas *Pool* memiliki 133 file, kelas *PepperFrog* memiliki 206 file, kelas *Boana* memiliki 172 file, kelas *South* memiliki 93 file, kelas *Dendropsophus* memiliki 132 file, kelas *Leptodactylus* memiliki 93 file, kelas *Rana* memiliki 426 file, kelas *Rhinella* memiliki 139 file, dan kelas *Scinax* memiliki 133 file. Tahap segmentasi ini bertujuan untuk menghasilkan data audio yang terstruktur dengan baik, siap untuk analisis lebih lanjut atau pelatihan model, serta memastikan kualitas data yang diperoleh tetap optimal.

2. Ekstraksi MFCC

Proses ekstraksi *Mel-frequency cepstral coefficients* (MFCC) bertujuan untuk mendapatkan fitur yang menggambarkan karakteristik suara, yang sering digunakan dalam analisis audio, seperti pengenalan suara dan klasifikasi audio. Langkah pertama adalah menentukan lokasi folder yang berisi file audio dan folder untuk menyimpan hasil ekstraksi. Setiap file audio dengan format wav diproses satu per satu.

Koefisien MFCC yang terdiri dari 13 komponen diekstraksi untuk mewakili frekuensi yang relevan dengan persepsi manusia terhadap suara, yang memungkinkan sistem untuk mengidentifikasi pola dalam sinyal audio. Hasil ekstraksi kemudian divisualisasikan untuk memberikan gambaran tentang bagaimana koefisien ini berubah seiring waktu, yang memudahkan pemahaman tentang karakteristik audio tersebut. Setelah itu, koefisien MFCC disimpan dalam format file yang memudahkan untuk analisis selanjutnya. Proses ini menghasilkan 1512 file MFCC yang dapat digunakan untuk analisis audio lebih lanjut.

Model 1 - Modified DenseNet 121

penelitian ini mengembangkan arsitektur Modified DenseNet-121, sebuah model jaringan saraf konvolusional yang didesain untuk meningkatkan akurasi dan efisiensi dalam klasifikasi dengan memanfaatkan fitur tambahan seperti normalisasi batch, dropout, dan pooling adaptif.

Basis Arsitektur

Arsitektur dasar yang digunakan adalah DenseNet-121. Model ini memiliki karakteristik konektivitas yang unik, di mana setiap lapisan menerima input dari semua lapisan sebelumnya. Dengan 121 lapisan konvolusional, konfigurasi ini memungkinkan aliran informasi yang lebih baik, sehingga meningkatkan kemampuan ekstraksi fitur.

Modifikasi Struktural

1. Normalisasi Batch (Batch Normalization)

Teknik ini diterapkan setelah tahap ekstraksi fitur utama DenseNet-121 dengan 1024 saluran untuk menstandarisasi distribusi aktivasi. Hal ini mengurangi pergeseran distribusi internal dan mempercepat konvergensi saat pelatihan.

2. Regularisasi (Dropout):

Untuk mencegah overfitting, dua lapisan dropout diintegrasikan ke dalam arsitektur model. Lapisan dropout pertama diterapkan setelah proses normalisasi batch, sementara lapisan dropout kedua ditempatkan sebelum lapisan klasifikasi. Strategi ini secara acak menonaktifkan sebagian neuron selama pelatihan, sehingga mendorong model untuk belajar representasi yang lebih robust dan meningkatkan kemampuan generalisasi terhadap data baru. Dengan pendekatan ini, model diharapkan dapat mengurangi ketergantungan terhadap fitur tertentu dan mengoptimalkan performanya pada tugas klasifikasi.

3. Reduksi Dimensi (Adaptive Average Pooling) :

Adaptive average pooling digunakan untuk mereduksi dimensi fitur menjadi ukuran tetap 1x1. Setelah itu, fitur diratakan (flatten) menjadi vektor satu dimensi sebagai input ke lapisan klasifikasi.

Lapisan Klasifikasi

Lapisan klasifikasi dimodifikasi dengan menggunakan fully connected layer yang dapat disesuaikan dengan jumlah kelas target. Dalam penelitian ini, model dirancang untuk mengklasifikasikan data ke dalam 9 kelas, memberikan fleksibilitas sesuai kebutuhan spesifik penelitian.

Strategi Optimasi

1. Optimizer:

Optimizer yang digunakan adalah Adam dengan learning rate sebesar $1e-4$ dan weight decay sebesar $1e-4$. Penggunaan learning rate

rendah dan regularisasi weight decay bertujuan untuk mencegah overfitting sekaligus mendukung konvergensi yang stabil selama proses pelatihan. Sebagai fungsi loss, digunakan CrossEntropyLoss untuk mengoptimalkan pembelajaran klasifikasi multi-kelas.

2. Penjadwal *Learning Rate*: model ini memanfaatkan ReduceLROnPlateau scheduler, yang secara adaptif menurunkan learning rate jika tidak ada penurunan pada nilai loss validasi setelah 3 epoch. Penjadwal ini dilengkapi dengan parameter verbose untuk memberikan laporan langsung selama pelatihan, sehingga memungkinkan pemantauan yang lebih mudah terhadap dinamika optimasi. Strategi ini dirancang untuk meningkatkan efisiensi dan stabilitas proses pelatihan.

Model 2- Modified DenseNet 169

Penelitian ini mengembangkan arsitektur Modified DenseNet169, sebuah model jaringan saraf konvolusional yang didesain secara komprehensif untuk meningkatkan kemampuan ekstraksi dan klasifikasi fitur.

Basis Arsitektur

Arsitektur dasar yang digunakan adalah DenseNet-169, yang dipilih karena karakteristik uniknya dalam konektivitas antar-lapisan.

Model ini memiliki 169 lapisan konvolusional dengan mekanisme koneksi yang memungkinkan setiap lapisan menerima input dari seluruh lapisan sebelumnya. Konfigurasi ini memfasilitasi aliran informasi yang optimal dan memungkinkan ekstraksi fitur yang lebih kompleks dibandingkan dengan arsitektur konvensional.

Modifikasi Struktural

1. Normalisasi Batch (Batch Normalization): Teknik ini diterapkan pada output DenseNet-169 dengan 1664 saluran untuk menstandarisasi distribusi aktivasi yang mengurangi pergeseran internal distribusi data dan mempercepat konvergensi selama proses pelatihan.
2. Regularisasi (Dropout): Dua lapisan dropout diintegrasikan ke dalam arsitektur. Lapisan pertama ditempatkan setelah batch normalization, sedangkan lapisan kedua diposisikan sebelum lapisan klasifikasi yang berfungsi mengurangi overfitting dengan cara menonaktifkan sebagian neuron secara acak selama proses pelatihan.
3. Reduksi Dimensi (Adaptive Average Pooling): Untuk mengurangi kompleksitas komputasional, adaptive average pooling digunakan guna mereduksi ukuran fitur menjadi resolusi tetap 1x1. Selanjutnya, fitur diratakan (flatten) menjadi vektor satu dimensi sebagai input lapisan klasifikasi.

Lapisan Klasifikasi

Lapisan klasifikasi dimodifikasi dengan menggunakan fully connected layer yang dapat disesuaikan dengan jumlah kelas target. Dalam penelitian ini, model dirancang untuk mengklasifikasikan data ke dalam 9 kelas, memberikan fleksibilitas sesuai kebutuhan spesifik penelitian.

Strategi Optimasi

1. *Optimizer*:

Adam *optimizer* dipilih dengan learning rate rendah ($1e-4$) dan *weight decay* ($1e-4$) untuk mencegah *overfitting* dan mendukung konvergensi yang stabil.

2. Penjadwal *Learning Rate*: ReduceLROnPlateau *scheduler* diimplementasikan untuk secara otomatis menurunkan learning rate jika tidak terdapat penurunan *loss* validasi setelah beberapa *epoch*, sehingga meningkatkan efisiensi pelatihan.

Hyperparameter

Konfigurasi hyperparameter pada arsitektur Modified DenseNet dirancang secara komprehensif untuk mengoptimalkan proses pembelajaran dan generalisasi model. Implementasi transfer learning dari ImageNet dikombinasikan dengan parameter struktural yang mencakup *batch normalization* (momentum 0.99, epsilon 1e-5) dan strategi *dropout* ganda (probabilitas 0.5) pada posisi strategis - setelah *batch normalization* dan sebelum lapisan klasifikasi akhir. Parameter optimasi menggunakan Adam *optimizer* dengan *learning rate* awal 1e-4, *weight decay* 1e-4, dan parameter beta ($\beta_1 = 0.9$, $\beta_2 = 0.999$, epsilon 1e-8). Pembelajaran adaptif ditingkatkan melalui implementasi ReduceLROnPlateau *scheduler* dengan mode 'min', faktor penurunan 0.1, *patience* 3 *epoch*, dan *learning rate minimum* 1e-6.

Pemilihan hyperparameter ini didasarkan pada keseimbangan antara kecepatan konvergensi, stabilitas pelatihan, dan kemampuan generalisasi. Batch normalization dengan momentum tinggi memastikan estimasi statistik yang stabil, sementara *dropout* ganda memberikan regularisasi yang efektif. Learning rate scheduler dengan *patience* moderat memungkinkan model mencapai konvergensi lokal sebelum penyesuaian, sementara faktor penurunan dan batas minimum learning rate mendukung *fine-tuning* yang efektif.

Metrik Evaluasi

Dalam mengevaluasi kinerja model, penelitian ini menggunakan beberapa metrik evaluasi yang umum digunakan dalam masalah klasifikasi. Metrik-metrik yang digunakan meliputi *akurasi*, *precision*, *recall*, dan *F1-score*. Perhitungan metrik-metrik klasifikasi ini didasarkan pada nilai-nilai yang diperoleh dari *confusion matrix*, yang terdiri dari *True Positive (TP)*, *False Negative (FN)*, *False Positive (FP)*, dan *True Negative (TN)* [8].

Confusion matrix memberikan gambaran komprehensif tentang performa model klasifikasi, di mana:

- *True Positive (TP)* menunjukkan jumlah prediksi benar pada kelas positif
- *False Negative (FN)* menunjukkan jumlah prediksi salah pada kelas negatif
- *False Positive (FP)* merepresentasikan jumlah prediksi salah pada kelas positif
- *True Negative (TN)* menggambarkan jumlah prediksi benar pada kelas negatif

$$Accuracy(\%) = \frac{TP + TN}{TP + FP + FN}$$

$$Precision(\%) = \frac{TP}{(FP + TP)}$$

$$Recall(\%) = \frac{TP}{(FN + TP)}$$

$$F1 - Score = \frac{(2 \times Recall \times Precision)}{(Recall + Precision)}$$

Gambar 1 memberikan ilustrasi visual tentang perhitungan *akurasi*, *precision*, dan *recall*(TRP)

berdasarkan nilai-nilai yang diperoleh dari *confusion matrix* tersebut.

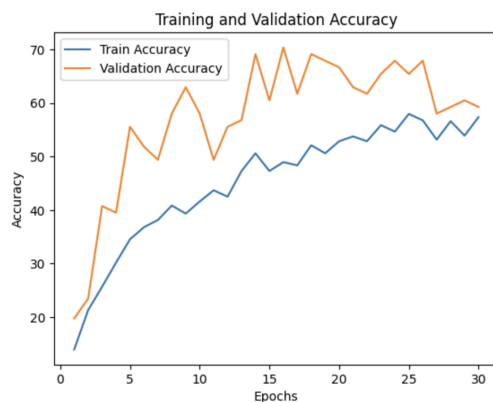
		Predicted class		
		Classified positive	Classified negative	
Actual class	Actual positive	TP	FN	TPR: $\frac{TP}{TP + FN}$
	Actual negative	FP	TN	FPR: $\frac{FN}{TN + FP}$
		Precision: $\frac{TP}{TP + FP}$	Accuracy: $\frac{TP + TN}{TP + TN + FP + FN}$	

Gambar 1. Ilustrasi confusion matrix [9]

Hasil dan Pembahasan

Pengujian Model 1 - Modified DenseNet 121

Pada tahap ini, dilakukan pengujian terhadap model DenseNet-121 untuk klasifikasi suara katak. Hasil pengujian dapat dilihat melalui grafik yang menunjukkan akurasi pelatihan dan validasi, yang menggambarkan peningkatan kinerja model selama proses pelatihan.

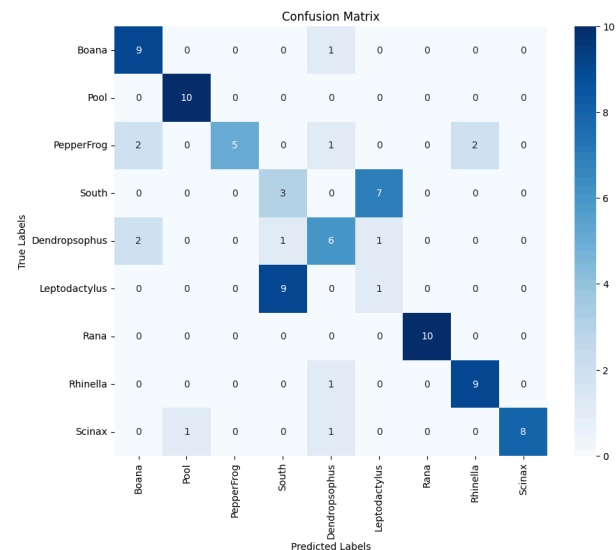


Gambar 2. Grafik Akurasi Model DenseNet-121

Berdasarkan hasil pengujian dari model modified DenseNet-121, grafik menunjukkan bahwa perkembangan pembelajaran yang baik selama 30 epoch pelatihan. Performa model diukur

menggunakan dua metrik utama, akurasi pelatihan (train accuracy) dan akurasi validasi (validation accuracy). Akurasi pelatihan meningkat secara bertahap dari 15% hingga 57%, sementara akurasi validasi mencapai puncaknya 70% pada epoch ke-15. Meskipun terdapat indikasi overfitting yang terlihat dari perbedaan antara akurasi pelatihan dan validasi, model tetap menunjukkan kemampuan generalisasi yang baik dengan akurasi validasi yang stabil di atas 60%. Secara keseluruhan, model DenseNet-121 menunjukkan performa yang baik dalam klasifikasi suara katak dengan akurasi yang memadai pada dataset validasi.

Analisis lebih lanjut dilakukan dengan evaluasi menggunakan confusion matrix untuk mengetahui secara rinci akurasi klasifikasi pada tingkat spesies. Hasil evaluasi ini membantu memvalidasi kinerja model yang sebelumnya terlihat dari grafik akurasi pelatihan dan validasi.



Gambar 3. Confusion Matriks model DenseNet-121

Matriks menunjukkan akurasi yang baik, dengan sebagian besar prediksi terkonsentrasi pada diagonal utama, yang menandakan klasifikasi yang tepat. Spesies Rana dan Pool mencapai akurasi sempurna dengan 10 prediksi yang benar, sementara Boana, Rhinella, dan Scinax memperoleh akurasi 90% dengan satu misklasifikasi. Spesies PepperFrog memiliki 5 prediksi tepat dari 10 sampel, dengan misklasifikasi terhadap Boana dan Rhinella. Misklasifikasi juga terjadi pada spesies Dendropsophus, dengan 6 prediksi tepat dari 10 sampel. Secara keseluruhan, confusion matrix menunjukkan bahwa model memiliki performa yang baik, dengan mayoritas spesies mencapai akurasi di atas 80%. Untuk mengukur performa model secara kuantitatif, dihitung metrik seperti akurasi, precision, recall, dan F1 score. Hasil evaluasi kuantitatif model DenseNet-121 dapat dilihat pada Tabel 2.

Tabel 2. Evaluasi Performa Model DenseNet-121

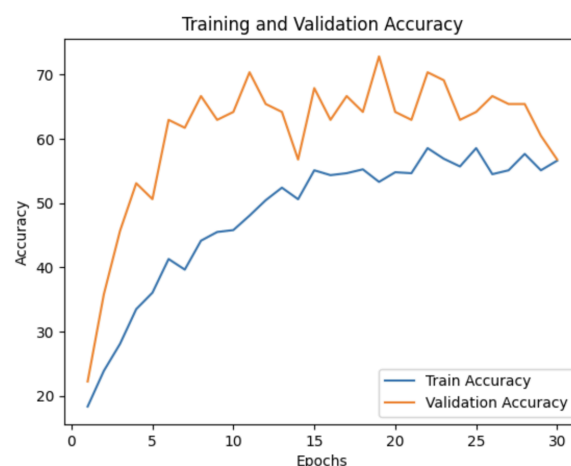
Metrik	Nilai (%)
Akurasi	68
Precision	71
Recall	68
F1 Score	68

Berdasarkan metrik evaluasi pada Tabel 2, model DenseNet-121 mencapai akurasi sebesar 68%, yang berarti 68% prediksi diklasifikasikan dengan benar. Precision sebesar 71% menunjukkan ketepatan model dalam memprediksi positif, di mana 71% dari prediksi positif adalah benar positif. Recall yang mencapai 68% menunjukkan kemampuan model dalam mendeteksi kasus positif, dengan 68% dari kasus positif berhasil

teridentifikasi. F1 Score sebesar 68% menggambarkan keseimbangan antara precision dan recall, menunjukkan performa model yang seimbang. Secara keseluruhan, metrik-metrik ini menunjukkan bahwa model memiliki performa yang konsisten dengan keseimbangan yang baik antara precision dan recall.

Pengujian Model 2 - Modified DenseNet 169

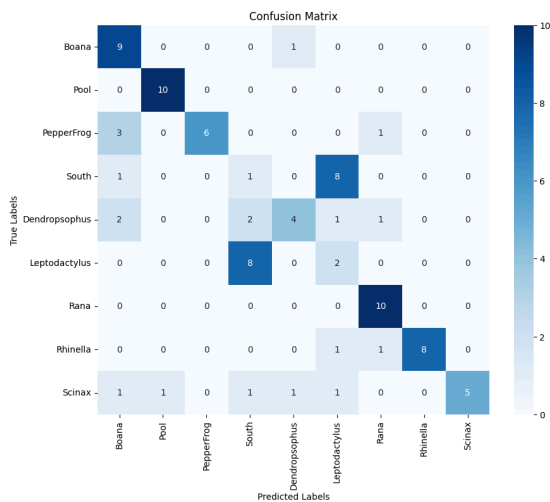
Pada tahap ini, dilakukan evaluasi menyeluruh terhadap model DenseNet-169 dalam klasifikasi suara katak. Evaluasi kinerja model dilakukan dengan menggunakan berbagai metrik dan visualisasi untuk memahami kemampuan model.



Gambar 4. Grafik Akurasi Model DenseNet-169

Berdasarkan hasil pengujian model DenseNet-169, grafik menunjukkan peningkatan pembelajaran yang konsisten selama 30 epoch pelatihan. Akurasi pada data latih meningkat dari 20% hingga 57%, sementara akurasi pada data validasi menunjukkan fluktuasi, dengan nilai tertinggi 70% pada epoch ke-20. Meskipun ada indikasi overfitting yang terlihat dari perbedaan

antara akurasi latih dan validasi, model tetap menunjukkan kemampuan generalisasi yang baik, dengan akurasi validasi yang stabil di atas 60%. Evaluasi lebih lanjut dilakukan dengan menggunakan confusion matrix untuk melihat lebih rinci akurasi klasifikasi pada tingkat spesies.



Gambar 5. Confusion Matriks model DenseNet-169

Matriks menunjukkan performa klasifikasi yang bervariasi antar spesies. Spesies Rana, Pool, dan Boana menunjukkan akurasi tinggi dengan masing-masing mencapai 10, 10, dan 9 prediksi yang benar. Sementara itu, beberapa spesies seperti Dendropsophus dan Scinax menghadapi tantangan klasifikasi yang lebih besar dengan tingkat akurasi yang lebih rendah, yang menunjukkan kompleksitas dalam membedakan karakteristik akustik tertentu. Evaluasi kuantitatif model disajikan melalui empat metrik utama yang dapat dilihat pada Tabel 3.

Tabel 3. Evaluasi Performa Model DenseNet-169

Metrik	Nilai (%)
Akurasi	61
Precision	68
Recall	61
F1 Score	62

Model mencapai akurasi global sebesar 61%, dengan precision yang lebih tinggi yaitu 68%. Recall model mencapai 61%, sementara F1 Score berada di angka 62%. Metrik-metrik ini menunjukkan bahwa model memiliki performa yang baik, dengan keseimbangan antara precision dan recall, meskipun masih ada ruang untuk perbaikan dalam akurasi klasifikasi secara keseluruhan. Model DenseNet-169 menunjukkan kemampuan yang baik dalam mengidentifikasi beberapa spesies katak, meskipun masih menghadapi kesulitan dalam membedakan karakteristik akustik yang lebih kompleks pada beberapa spesies lainnya.

Kesimpulan

Berdasarkan hasil klasifikasi suara katak, Model Modified DenseNet-121 mencapai akurasi 68% dan memiliki kinerja lebih baik dibandingkan Modified DenseNet-169, dengan precision, recall, dan F1-score yang lebih tinggi. Hasil ini menunjukkan bahwa deep learning dapat digunakan untuk mendukung pemantauan keanekaragaman hayati dan konservasi katak. Penelitian selanjutnya disarankan untuk mengembangkan model yang lebih baik dan menggunakan dataset yang lebih besar serta beragam guna meningkatkan akurasi dan generalisasi.

Ucapan Terima Kasih

Ucapan terima kasih kepada Bapak Ardika Satria, M. Si dan tim pengajar *Deep Learning* Sains Data ITERA, yang telah memberikan arahan dan bimbingan hingga selesainya penulisan paper ini. Penelitian ini diharapkan dapat membantu upaya pemantauan keanekaragaman hayati dan konservasi spesies dengan memanfaatkan teknologi *deep learning*. Penelitian selanjutnya dapat mengoptimalkan model *Modified Desnet-121* dan *DesNet-169* dengan memodifikasi arsitektur untuk meningkatkan akurasi model.

- [12] M. Richter, M. Kurpas, and M. Maska, "Learning by confusion approach to characterize phase transitions," arXiv:2206.15114, Jun. 2022, doi: 10.48550/arXiv.2206.15114.

References

- [1] D. Iskandar, *The Amphibians of Java and Bali*. Research and Development Centre for Biology LIPI, 1998.
- [2] M. A. Rofiq, U. Usman, and I. Wahyuni, "Keanekaragaman Amfibi (Ordo Anura) Berdasarkan Tipe Habitat di Taman Wisata Alam Pulau Sangiang," *Prosiding Seminar Nasional Biologi*, vol. 9, pp. 202–213, 2021.
- [3] Dedieko, "Save The Frog Day: Katak, Bioindikator Perubahan Lingkungan," *Universitas Gajah Mada Fakultas Biologi*, 2013. [Online]. Available: <https://biologi.ugm.ac.id/2013/05/07/save-the-frog-day-katak-bioindikator-perubahan-lingkungan/>.
- [4] M. Nisa and K. Kismiyati, "Cultivation technique of *Chanos chanos* modular system and semi-intensive at the center for brackish water aquaculture (BBPBAP) Jepara, Central of Java," *IOP Conference Series: Earth and Environmental Science*, vol. 441, no. 1, 2020.
- [5] M. D. Kusriani, "Konservasi Amfibi Di Indonesia: Masalah Global Dan Tantangan," *Media Konservasi*, vol. 12, no. 2, pp. 89–95, 2007.
- [6] R. Umar, I. Riadi, and A. Hanif, "Analisis Bentuk Pola Suara Menggunakan Ekstraksi Ciri Mel-Frequency Cepstral Coefficients (MFCC)," *Cogito Smart Journal*, vol. 4, no. 2, pp. 294–304, 2018.
- [7] W. S. Mada Sanjaya and Z. Salleh, "Implementasi Pengenalan Pola Suara Menggunakan Mel-Frequency Cepstrum Coefficients (MFCC) dan Adaptive Neuro-Fuzzy Inference System (ANFIS) Sebagai Kontrol Lampu Otomatis," *Al-Hazen Journal of Physics*, 2014.
- [8] Z. J. Ruff, D. B. Lesmeister, C. L. Appel, and C. M. Sullivan, "Workflow and convolutional neural network for automated identification of animal sounds," *Ecological Indicators*, vol. 124, Art. no. 107419, 2021.
- [9] S. Wei, S. Zou, F. Liao, and others, "A comparison on data augmentation methods based on deep learning for audio classification," in *Journal of Physics: Conference Series*, vol. 1453, no. 1, pp. 012085, 2020, IOP Publishing.
- [10] L. Nanni, G. Maguolo, S. Brahnam, and M. Paci, "An ensemble of convolutional neural networks for audio classification," *Applied Sciences*, vol. 11, no. 13, p. 5796, 2021, MDPI.
- [11] Rendi Nurcahyo and Mohammad Iqbal, "Pengenalan Emosi Pembicara Menggunakan Convolutional Neural Networks", *J. RESTI (Rekayasa Sist. Teknol. Inf.)*, vol. 6, no. 1, pp. 115 - 122, Feb. 2022.