

# Klasifikasi Suara Kucing dan Anjing Menggunakan *Convolutional Neural Network*

Cecilia Tania Emanuella<sup>1</sup>, Musfita<sup>2,3</sup>, dan Armin Lawi<sup>3</sup>

Program Studi Sistem Informasi, Departemen Matematika,

Fakultas MIPA, Universitas Hasanuddin

[emanuella18h@student.unhas.ac.id](mailto:emanuella18h@student.unhas.ac.id), [musfita18h@student.unhas.ac.id](mailto:musfita18h@student.unhas.ac.id), [armin@unhas.ac.id](mailto:armin@unhas.ac.id)

**Abstract** — Cats and dogs are human pets that communicate through sound. To identify, distinguish and recognize the voices of other pets, sound classification is needed to make it clearer. This study presents a classification method using Deep Learning techniques with a Convolutional Neural Network (CNN) model on a limited amount of data. This classification requires class labels and the same data size by performing a generation that takes the same sound length in 2 classes, then MFCC feature extraction is carried out to obtain information in the audio data. CNN is composed of 3 convolutions with filters 16, 32, 64 and kernel 3x3, max pooling 2x2, dropout, flatten and 3 hidden layers, and neurons 128, 64, 2 with relu and softmax activation functions. After the data is shared for training and testing then create a CNN model and train the data and predict the model. The best training accuracy results are 93.8% and the highest validation accuracy is 90%, for the last epoch the training accuracy is 92.2% and the validation accuracy is 87.7% with an f1-score of 88%. For precision 87% for cats and 89% for dogs and for recall for cats 89% and dogs for 87%, where the positive label is cats and the AUROC score is 87.8%.

**Keyword** — Classification, Convolution Neural Network (CNN), Data Pre-processing, Deep Learning, Feature Extraction.

**Abstrak** — Kucing dan anjing merupakan hewan peliharaan manusia yang berkomunikasi melalui suara. Untuk mengetahui, membedakan dan mengenali suara hewan peliharaan yang lain, dibutuhkan klasifikasi suara agar lebih jelas. Penelitian ini menyajikan metode klasifikasi menggunakan teknik *Deep Learning* dengan model *Convolutional Neural Network* (CNN) pada data dengan jumlah yang terbatas. Klasifikasi ini memerlukan label kelas dan ukuran data yang sama dengan melakukan generation yang mengambil panjang suara yang sama pada 2 kelas, kemudian dilakukan ekstraksi fitur MFCC mendapatkan informasi dalam data audio. CNN tersusun dari 3 *convolution* dengan filter 16, 32, 64 dan masing-masing kernel 3x3, *max pooling* 2x2, *dropout*, *flatten* dan 3 *hidden layers*, serta neuron 128, 64, 2 dengan fungsi aktivasi *relu* dan *softmax*. Setelah data dibagi untuk *training* dan *testing* kemudian membuat model CNN dan melatih data serta memprediksi model. Hasil akurasi latih terbaik sebesar 93.8% dan akurasi validasi tertinggi 90%, untuk epoch terakhir akurasi latih 92.2% dan akurasi validasi 87.7% dengan *f1-score* 88%. Untuk *precision* kucing 87% dan anjing 89% serta *recall* kucing 89% dan anjing 87%, dimana label positifnya yaitu kucing dan skor AUROC 87.8%.

**Kata kunci** — *Convolution Neural Network* (CNN), *Deep Learning*, Ekstraksi Fitur, Klasifikasi, Pra-pengolahan Data

## I. PENDAHULUAN

Kucing dan anjing merupakan salah satu hewan peliharaan manusia yang berkomunikasi melalui suara. Kucing dan anjing merupakan salah satu hewan peliharaan manusia yang berkomunikasi melalui suara. Untuk membantu mengetahui, membedakan dan mengenali suara hewan dengan peliharaan yang lain, dibutuhkan klasifikasi suara hewan agar lebih jelas.

Generasi suara dan sistem persepsi hewan telah berevolusi untuk membantu mereka bertahan hidup di lingkungan mereka. Dari perspektif evolusi, suara yang disengaja yang dihasilkan oleh hewan harus berbeda dari suara acak lingkungan. Beberapa hewan memiliki keistimewaan kemampuan sensorik, seperti penglihatan, pemandangan, perasaan, dan kesadaran akan perubahan alam yang dibandingkan kepada manusia. Suara binatang dapat bermanfaat bagi manusia dalam hal keamanan, prediksi bencana alam, dan interaksi intim jika kita mampu mengenalinya dengan baik[1].

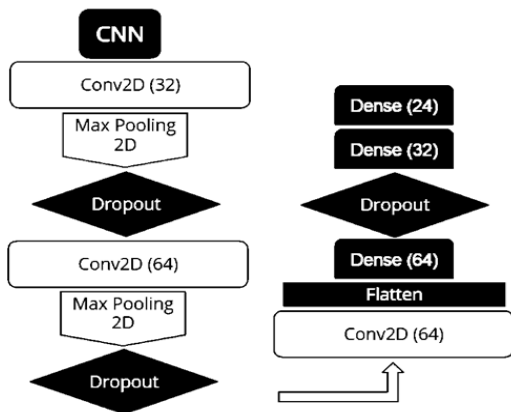
Peneliti dalam beberapa tahun terakhir, dan beberapa penelitian telah dilakukan pada klasifikasi suara hewan[1]-[4]-[5]-[9] dimana pada penelitian mereka melakukan klasifikasi suara hewan dengan menggunakan metode dan model yang berbeda, serta dengan objek yang berbeda pula untuk menghasilkan penelitian yang sesuai yang diharapkan, dengan akurasi yang cukup baik. Dalam penelitian kali ini akan dilakukan klasifikasi suara hewan kucing dan anjing untuk membedakan suara di antara kedua hewan tersebut. Sebelumnya sudah ada penelitian yang sama tentang klasifikasi perbedaan kucing dan anjing melalui perbedaan gambar pada penelitian[9].

Yang membedakan penelitian sebelumnya dengan penelitian yang akan kami lakukan yaitu melakukan klasifikasi suara kucing dan anjing dengan menggunakan data yang tersedia dalam jumlah yang terbatas. Metode klasifikasi yang digunakan adalah teknik *Deep Learning* dengan model CNN. *Convolutional Neural Network* adalah salah satu metode machine learning dari pengembangan *Multi Layer Perceptron* (MLP) yang didesain untuk mengolah data dua dimensi. CNN tersusun dari 3 *convolution* dengan filter 16, 32, 64 dan masing-masing kernel 3x3, *max pooling* 2x2, *dropout*, *flatten* dan 3 *hidden layer*, serta neuron 128, 64, 2 dengan fungsi aktivasi *relu* dan *softmax*.

Untuk lebih memahami hasil yang didapatkan, disajikan *confusion matrix*, kurva ROC dan *report classification*. Model CNN untuk klasifikasi memerlukan label kelas dan ukuran data yang sama sehingga kumpulan data (dataset) dibangun dengan melakukan generation pada data dengan mengambil panjang suara yang sama pada 2 kelas, kemudian dilakukan ekstraksi fitur dengan MFCC untuk membantu mendapatkan banyak informasi dalam data audio. Setelah data dibagi untuk training dan validasi kemudian membuat model CNN dan melatih data serta memprediksi model.

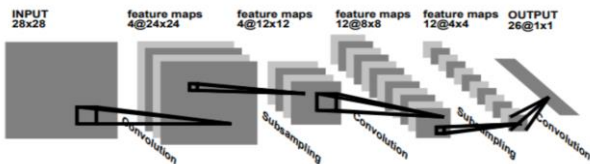
## II. TINJAUAN PUSTAKA

### A. Konsep Dasar



Gambar 1. Sekuens Convolutional Neural Network.

Convolutional Neural Network adalah salah satu metode machine learning dari pengembangan *Multi Layer Perceptron* (MLP) yang didesain untuk mengolah data dua dimensi. CNN termasuk dalam jenis *Deep Neural Network* karena dalamnya tingkat jaringan dan banyak diimplementasikan dalam data citra. CNN dirancang untuk memproses suatu data yang ada dalam bentuk banyak array, contohnya gambar warna yang terdiri dari 2D array yang mengandung piksel dalam tiga macam warna yaitu Red, Green, dan Blue. Ada berbagai macam bentuk CNN adalah 1D untuk sinyal dan urutan biasa nya digunakan untuk bahasa, 2D untuk gambar atau suara; dan 3D untuk video atau gambar volumetrik.



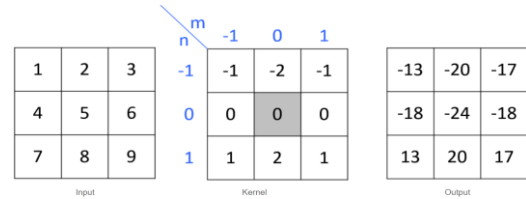
Gambar 2. Convolutional Neural Network.

Dalam gambar diatas, dijelaskan bahwa cara kerja *Convolution Neural Network* adalah dengan memberlakukan filter pada suatu gambar yang dijalankan dengan menggunakan matriks tersebut[4]. Kemudian hasil filtering tersebut difilter kembali untuk mendapatkan suatu

subsampling[5]. Kemudian sub sampling tersebut diolah ke dalam suatu pemetaan fitur. Hasil dari pemetaan tersebut diklasifikasikan ke dalam pilihan solusi yang tersedia[6].

### B. Komponen Arsitektur dan Klasifikasi

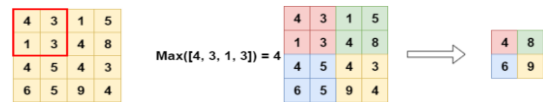
#### 1) Convolution 2D



Gambar 3. Convolutional 2D.

Konvolusi adalah jenis operasi matriks, yang terdiri dari kernel, matriks kecil bobot, yang meluncur di atas data masukan yang melakukan perkalian bijak dengan bagian masukannya, lalu menjumlahkan hasilnya menjadi keluaran. Matriks direferensikan disini sebagai [kolom, baris], bukan [baris, kolom]. M adalah arah horizontal( kolom) dan N adalah arah vertikal (baris).

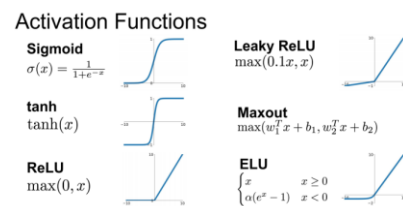
#### 2) Max Pooling 2D



Gambar 4. Max Pooling 2D.

*Max Pooling* adalah operasi penggabungan yang menghitung nilai maksimum untuk patch peta fitur, dan menggunakannya untuk membuat peta fitur downsamped. Digunakan setelah lapisan konvolusional.

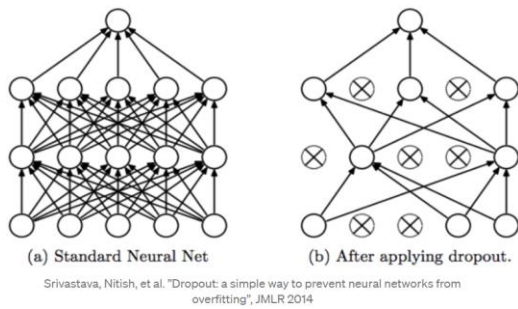
#### 3) Activation ReLU



Gambar 5. Activation.

Activation ReLU (*Rectified Linear Units*) adalah jenis fungsi aktivasi yang linear di dimensi positif, tetapi nol di dimensi negatif. Ketegaran dalam fungsi adalah sumber non-linearitas. Linearitas dalam dimensi positif memiliki sifat menarik yang mencegah gradien non-saturasi (kontras dengan aktivasi sigmoid), meskipun untuk setengah dari garis nyata gradiennya adalah nol.

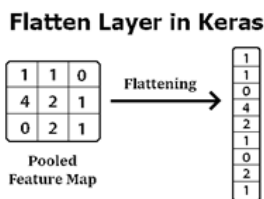
4) Dropout



Gambar 6. Droupout.

Dropout mengacu pada pengabaian neuron selama fase pelatihan dari rangkaian neuron tertentu yang dipilih secara acak, dropout digunakan untuk mencegah *overfitting*. Dropout mengubah konsep dari mempelajari semua bobot bersama-sama menjadi mempelajari sebagian kecil bobot dalam *network* di setiap iterasi pelatihan.

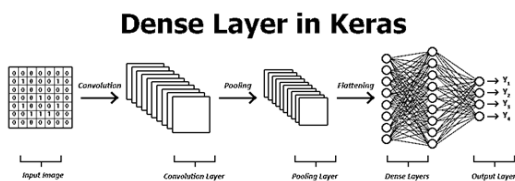
5) Flatten



Gambar 7. Flatten.

Flatten merupakan salah satu operasi pada tensor dengan membentuk kembali tensor menjadi bentuk yang sama dengan jumlah elemen yang terdapat dalam tensor. Ini sama dengan array element 1d.

6) Dense



Gambar 8. Dense.

Dense merupakan lapisan *layer* pada jaringan saraf yang terhubung secara dalam, artinya setiap neuron pada lapisan padat tersebut menerima masukan dari semua neuron pada lapisan sebelumnya. Dense merupakan lapisan yang paling umum digunakan dalam model.

7) Data Generation

Semua audio memiliki panjang yang berbeda sehingga menjadi masalah karena dalam pengklasifikasian hanya menerima data dengan

bentuk yang sama untuk setiap sampel. Sehingga dalam penelitian ini menggunakan data audio 2 detik dengan metode pengambilan acak dari semua sampel.

8) Feature Extraction

Raw audio wave yang diekstrak menggunakan librosa tidak terlalu informatif karena pada dasarnya hanya terdiri dari data satu dimensi yang disimpan dalam array yang hanya mewakili amplitudo setiap bit. Namun, kita perlu juga untuk mempertimbangkan pitch audio tersebut sehingga untuk mengekstrak informasi nada berdasar audio yang diberikan digunakan fungsi mfcc(). Mel Frequency Cepstral Coefficients (MFCC) merupakan metode ekstraksi fitur untuk mendapatkan lebih banyak informasi dalam data audio. MFCC sebagai vektor ciri yang baik untuk merepresentasikan suara dan sinyal yang telah terbukti bermanfaat untuk pengenalan suara, MFCC dapat digunakan untuk mengekstrak data sinyal suara sehingga bisa didapatkan ciri yang terdapat pada data sinyal suara. Pada MFCC diterapkan transpose dan output dari fitur ini ditunjukkan dengan gambar heatmap yang nantinya akan dilakukan klasifikasi pada gambar tersebut.

9) Label Preprocessing

Sebelum membangun arsitektur kira perlu memberi *label-encode* dan *one-hot-encode* dari setiap sampel dengan representasi 0 dan 1 karena nilai dalam array *generate* masih dalam bentuk data kategorikal yang sama sekali tidak dapat diolah untuk *neural network*.

10) Normalization

Normalisasi data yang kami gunakan yaitu *Min-Max Normalization*. Cara kerjanya setiap nilai pada sebuah fitur dikurangi dengan nilai minimum fitur tersebut, kemudian dibagi dengan rentang nilai atau nilai maksimum dikurangi nilai minimum dari fitur tersebut.

$$x_{new} = \frac{x_{old} - x_{min}}{x_{max} - x_{min}} \quad (1)$$

11) Data Split

Tahap pasca persiapan, data dibagi menjadi dataset train dan test menggunakan *Hold-Out* dengan pembagian masing-masing 70% dan 30%. Model CNN yang dalam kemudian dilatih pada data *train* dan memprediksi model dengan diuji pada data test.

12) Evaluation

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} \quad (4)$$

Predicted -- Positive/Negative	Actual -- True/False	
	True Positive	False Positive
False Negative	True Negative	

Dalam hasil pelatihan dilakukan pengujian seperti menguji *precision*, *accuracy*, *f1-score*, *loss function* dan *confusion matrix*. Dengan matriks menampilkan tabel hasil prediksi pada data *test*, *f1-score* sebagai poin data yang diprediksi sebagai persentase dari semua prediksi, *precision* mengukur seberapa akurat prediksi tersebut, *loss function* untuk mengevaluasi cara algoritma memodelkan data.

### C. Penelitian yang Relevan

(Lee et al., 2017) Pada penelitian ini dilakukan Klasifikasi Audio untuk Pengenalan audio-visual (AVR) Menggunakan Arsitektur CNN tingkat sampel. Dimana dalam penelitiannya mengusulkan penggunaan arsitektur jaringan saraf konvolusional 3D (3D CNN). Dengan menggunakan arsitektur jaringan yang relatif kecil dan kumpulan data yang jauh lebih kecil untuk pelatihan, metode yang kami usulkan melampaui kinerja metode serupa yang ada untuk pencocokan audio-visual, yang menggunakan CNN 3D untuk representasi fitur. Metode yang diusulkan mencapai peningkatan relatif lebih dari 20% pada tingkat kesalahan yang sama dan lebih dari 7% pada presisi rata-rata dibandingkan dengan metode mutakhir[2].

(Torfi et al., 2017) Penelitian Investigasi Berbagai Model Berbasis CNN untuk Klasifikasi Suara Burung yang Lebih Baik juga dilakukan oleh trofi. Dimana dalam studi ini, membandingkan yang berbeda model klasifikasi dan secara selektif memadukannya untuk lebih meningkatkan kinerja klasifikasi suara burung. Tiga jenis frekuensi waktu Representasi (TFR) suara burung diselidiki dengan tujuan untuk mengkarakterisasi komponen akustik yang berbeda burung: Spektogram mel, spektogram berbasis komponen harmonik, dan spesifikasi berbasis komponen perkusi trogram. Selain TFR yang berbeda, arsitektur pembelajaran mendalam yang berbeda, Sub Spectral Net, digunakan untuk mengklasifikasikan suara burung. Hasil percobaan pada pengklasifikasian 43 jenis burung menunjukkan bahwa fusing diseleksi dalam model pembelajaran secara efektif dapat meningkatkan kinerja klasifikasi. Model gabungan terbaik kami dapat mencapai a akurasi seimbang 86,31% dan skor F1 tertimbang 93,31%[4].

(Liu, Liu and Zhou, 2018) Melakukan penelitian terkait Klasifikasi gambar pada Anjing dan Kucing. Dalam penelitian, mereka mengembangkan algoritma untuk mengklasifikasikan gambar anjing dan kucing, yaitu

kompetisi Anjing vs Kucing dari Kaggle. Terutama menyelidiki dua pendekatan untuk mengatasi masalah ini. Yang pertama adalah model pengenalan pola tradisional. Kami mengekstrak beberapa fitur buatan manusia seperti warna dan Dense-SIFT, merepresentasikan gambar menggunakan model *bag of words*, lalu melatih pengklasifikasi *Support Vector Machines* (SVM). Untuk pendekatan kedua, kami menggunakan *Deep Convolutional Neural Networks* (CNN) untuk mempelajari fitur gambar dan Jaringan Syaraf Tiruan *Backpropagation* (BP) terlatih dan SVM untuk klasifikasi. Kami mencoba berbagai eksperimen untuk meningkatkan performa kami pada set data pengujian, dan akhirnya mendapatkan akurasi terbaik 94.00% dengan pendekatan kedua[9].

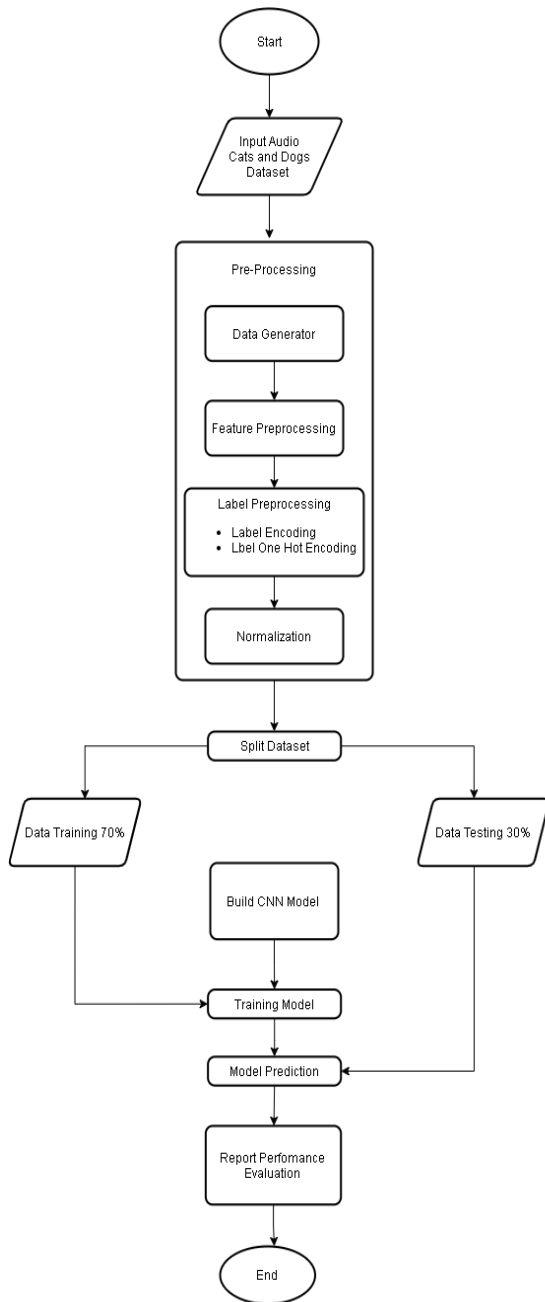
(Pandeya, Kim and Lee, 2018) melakukan penelitian tentang Klasifikasi suara kucing domestik menggunakan *Learned Features* dari *Deep Neural Nets*. Dimana dalam penelitiannya menggunakan dua jenis fitur yang dipelajari dari *deep neural networks*; satu dari *convolutional neural net* (CNN) pada data musik dengan *transfer learning* dan lainnya dari *unsupervised convolutional deep belief network* yaitu (CDBN) hanya dilatih pada kumpulan suara kucing. Dengan membandingkan kinerja klasifikasi faktor-faktor. Dimana jumlah data yang ditingkatkan dengan augmentasi, fitur yang dipelajari dari CNN terlatih atau CDBN tanpa pengawasan, GAP atau FDAP konvensional, dan pembelajaran mesin algoritma yang digunakan untuk klasifikasi. Selain itu, fitur yang dipelajari dari CNN terlatih dan produksi CDBN tanpa pengawasan hasil yang baik dalam percobaan. Oleh karena itu, dengan kombinasi dari semua faktor positif tersebut, diperoleh hasil terbaik dengan akurasi 91,13%, *f1-score* 0,91, dan area bawah kurva 0,995.(AUC)[1].

(Singh, 2020) Dalam penelitian Sing mengenai Klasifikasi Suara Hewan Menggunakan *Convolutional Neural Network* mengusulkan kerangka kerja berbasis jaringan saraf konvolusional yang fleksibel untuk klasifikasi audio hewan. Model didorong oleh ideologi untuk mengidentifikasi suara binatang dalam file audio dengan memaksa jaringan untuk memperhatikan efek audio inti yang ada dalam audio untuk menghasilkan *Mel-spectrogram*. Kerangka yang dirancang mencapai akurasi 98% saat mengklasifikasikan audio hewan pada set data berlabel mingguan. Teknologi mutakhir dalam penelitian ini adalah membangun kerangka kerja yang bahkan dapat berjalan pada mesin dasar dan tidak memerlukan perangkat kelas atas untuk menjalankan klasifikasi[5].

(Maccagno et al., 2021) Melakukan penelitian mengenai Klasifikasi audio di situs konstruksi dengan arsitektur yang sama dalam penelitian ini yaitu CNN. Dalam penelitian mereka mendemonstrasikan bahwa dimungkinkan untuk menerapkan pendekatan saraf sudah diuji dalam klasifikasi suara lingkungan ke yang lebih spesifik dan tantangan-domain lenge, yaitu salah satu lokasi konstruksi, memperoleh hasil yang lumayan tinggi. Sampai saat ini, arsitektur yang diusulkan telah diujicobakan pada lima kelas perolehan akurasi 97%[7].

### III. METODE PENELITIAN

#### A. Alur Kerja



Gambar 9. Alur Kerja.

#### B. Lokasi dan Waktu

Lokasi Penelitian dilakukan di Kota Makassar dan Maros pada tanggal 17 Mei 2021 - Selesai. Penelitian dilakukan di lokasi tersebut untuk melakukan klasifikasi pada dataset Audio *Cats and Dogs* dengan menggunakan CNN Pada waktu yang telah ditentukan.

TABEL 1  
Program Kerja

PROGRAM KERJA	PELAKSANAAN			
	MEI 2021			
	I	II	III	IV
Studi Literatur				
Melakukan Pengimputan Dataset Audio Cats and Dogs				
Melakukan Orientasi Data				
Melakukan Pre-Processing Data				
Melakukan Melakukan visualisasi data				
Membagi data dengan melakukan split data				
Membuat model CNN				
Training Model				
Prediksi Model				
Report Performance evaluation				

#### C. Sumber Data

Dalam penelitian ini dataset yang kami ambil berasal dari <https://www.kaggle.com/mmoreaux/audio-cats-and-dogs>. Dataset ini merupakan dataset yang berisi suara kucing dan anjing yang terdiri dari 164 file WAV suara kucing dengan audio 1323 detik serta 113 file WAV suara anjing dengan audio 598 detik.

TABEL 2  
Informasi Dataset

No	Data	Jumlah Data
1	<i>Cats</i>	164
2	<i>Dogs</i>	113

#### C. Instrumen Penelitian

Instrumen Penelitian Beberapa perangkat yang digunakan untuk mengerjakan tugas ini adalah sebagai berikut :

##### 1) Perangkat Lunak

Google Collab

Google Collaboratory atau Google Collab merupakan *tools* yang berbasis *cloud* dan *free* untuk tujuan penelitian. Google Collab dibuat dengan *environment jupyter* dan mendukung hampir semua *library* yang dibutuhkan untuk klasifikasi pada penelitian ini.

##### 2) Perangkat Keras

Laptop Asus dengan spesifikasi:

Intel(R) Core(TM) i7-8565U CPU @ 1.80GHz  
1.99 GHz

RAM 8 GB

64 bit *operating system*

Perangkat keras ini digunakan untuk mencari sebuah dataset untuk diolah dan dilakukan klasifikasi pada penelitian ini.

## IV. HASIL DAN PEMBAHASAN

### A. Model Klasifikasi

Membuat pengklasifikasi dengan model CNN di mana dalam penelitian ini menggunakan gambar *heatmap* seperti yang ada ditampilkan sebelumnya. Dengan fungsi *summary()* bisa dilihat detail dari arsitektur.

```

Constructing the neural network architecture
model = Sequential()
model.add(Conv2D(16, (3, 3), activation='relu', strides=(1, 1), padding='same', input_shape=input_shape))
model.add(MaxPooling2D((2, 2)))
model.add(Conv2D(32, (3, 3), activation='relu', strides=(1, 1), padding='same'))
model.add(MaxPooling2D((2, 2)))
model.add(Conv2D(64, (3, 3), activation='relu', strides=(1, 1), padding='same'))
model.add(MaxPooling2D((2, 2)))
model.add(Conv2D(128, (3, 3), activation='relu', strides=(1, 1), padding='same'))
model.add(MaxPooling2D((2, 2)))
model.add(Flatten())
model.add(Dense(128, activation='relu'))
model.add(Dropout(0.3))
model.add(Dense(64, activation='relu'))
model.add(Dropout(0.3))
model.add(Dense(2, activation='softmax'))

model.compile(loss='binary_crossentropy', optimizer='adam', metrics=['acc'])

model.summary()

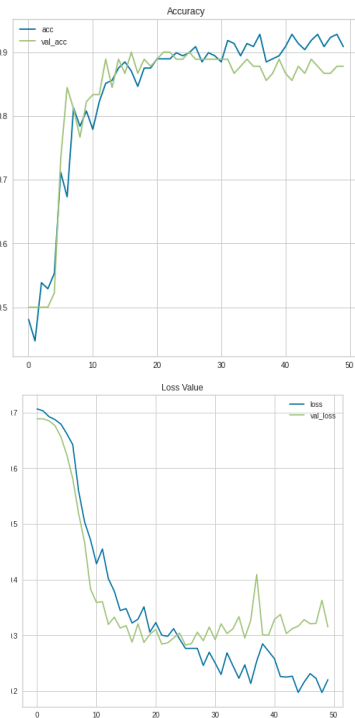
Model: "sequential"
Layer (type) Output Shape Param #
-----
conv2d (Conv2D) (None, 276, 13, 16) 160
max_pooling2d (MaxPooling2D) (None, 138, 6, 16) 0
conv2d_1 (Conv2D) (None, 138, 6, 32) 4640
max_pooling2d_1 (MaxPooling2D) (None, 69, 3, 32) 0
conv2d_2 (Conv2D) (None, 69, 3, 64) 18496
max_pooling2d_2 (MaxPooling2D) (None, 34, 1, 64) 0
dropout (Dropout) (None, 34, 1, 64) 0
flatten (Flatten) (None, 2174) 0
dense (Dense) (None, 128) 27856
dropout_1 (Dropout) (None, 128) 0
dense_1 (Dense) (None, 64) 8256
dropout_2 (Dropout) (None, 64) 0
dense_2 (Dense) (None, 2) 130
-----
total params: 318,338
trainable params: 318,338
non-trainable params: 0
    
```

Gambar 10. Membangun Model CNN.

### B. Evaluasi Kinerja

#### 1) Fitting Data

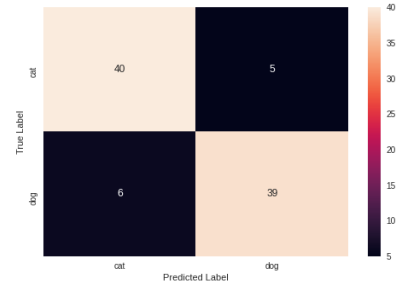
Setelah meng *compile* model, kemudian data dilatih dengan menggunakan 50 *epochs* dan diperoleh hasil plotting .



Gambar 11. Plot Hasil Fitting Data.

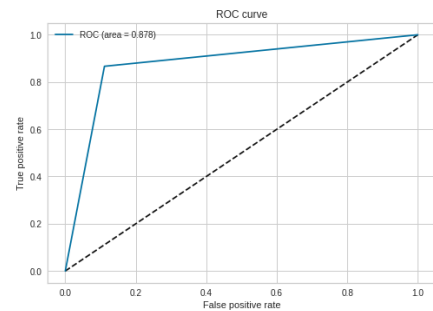
#### 2) Hasil Klasifikasi

Membuat *confusion matrix* untuk mengevaluasi performa model. Dari gambar terlihat ada yang salah klasifikasi, 5 data *cat* yang diprediksi sebagai *dog* dan 6 data *dog* diprediksi sebagai *cat*.



Gambar 12. Confusion Matrix.

Kemudian divisualisasikan kurva ROC AUC dari hasil klasifikasi di bawah ini.



Gambar 13. ROC Curve.

Berikut hasil *report* klasifikasi yang menampilkan *precision*, *recall*, *f1-score*, *support*, *accuracy*, *macro avg*, dan *weight avg*.

Classification Report				
	precision	recall	f1-score	support
0	0.87	0.89	0.88	45
1	0.89	0.87	0.88	45
accuracy			0.88	90
macro avg	0.88	0.88	0.88	90
weighted avg	0.88	0.88	0.88	90

Gambar 14. Classification Report.

### C. Pembahasan

Dari hasil penelitian dilakukan pada Dataset Audio *Cat and Dog* yang berisi tentang suara hewan pada kucing dan anjing. Dataset ini berjumlah 277 wav, yang terbagi atas 2 kelas yaitu kucing dan anjing. Upaya klasifikasi yang dilakukan metode klasifikasi dengan menerapkan arsitektur CNN. Pada penelitian ini dilakukan teknik pra-proses, *generate* data, ekstraksi fitur dan klasifikasi audio melalui hasil ekstraksi fitur gambar *heatmap*. Dilakukan klasifikasi ini, untuk dapat mengenali suara hewan pada kucing dan anjing dengan hasil yang cukup baik dengan proses *training* dan *testing* yang berbanding 70:30. Dengan hal itu, didapatkan evaluasi model dengan akurasi tertinggi 93,8

untuk validasi tertinggi 90 serta untuk *epoch* terakhir dengan akurasi 92,2 dan validasi 87,7 untuk *ROC curve* yang didapatkan yaitu 87,8 dan untuk masing-masing kelas terklasifikasi dengan baik.

## V. KESIMPULAN

Metode praproses dan metode klasifikasi dengan menggunakan *Convolutional Neural Network* cukup baik untuk menentukan kebenaran dari klasifikasi data audio. Hal ini terbukti dengan hasil akurasi sebesar 88%. Perubahan tingkat *confusion* tidak mempengaruhi hasil akurasi. Hal ini membuktikan bahwa klasifikasi menggunakan metode CNN relatif baik terhadap perubahan parameter yang dilakukan. Dengan menggunakan data training yang baik dan optimal, maka subset dari data *training* tersebut juga akan menghasilkan klasifikasi yang baik.

## UCAPAN TERIMA KASIH

Para penulis mengucapkan terima kasih kepada Tuhan Yang Maha Esa yang telah melimpahkan rahmat-Nya sehingga penulis dapat menyelesaikan penelitian ini. Penulis juga mengucapkan terima kasih kepada Bapak Armin Lawi dan Khawaritzmi Abdallah Ahmad yang telah membimbing penulis dalam mengerjakan penelitian ini.

## DAFTAR PUSTAKA

- [1] Y. R. Pandeya, D. Kim, and J. Lee, "Domestic cat sound classification using learned features from deep neural nets," *Appl. Sci.*, vol. 8, no. 10, pp. 1–17, 2018.
- [2] J. Lee, T. Kim, J. Park, and J. Nam, "Raw Waveform-based Audio Classification Using Sample-level CNN Architectures," no. Nips, 2017.
- [3] A. Torfi, S. M. Iranmanesh, N. Nasrabadi, and J. Dawson, "3D convolutional neural networks for cross audio-visual matching recognition," *IEEE Access*, vol. 5, pp. 22081–22091, 2017.
- [4] J. Xie, K. Hu, M. Zhu, J. Yu, and Q. Zhu, "Investigation of Different CNN-Based Models for Improved Bird Sound Classification," *IEEE Access*, vol. 7, pp. 175353–175361, 2019.
- [5] N. Singh, "Classification of Animal Sound Using Convolutional Neural Classification of Animal Sound Using Convolutional Neural Network Network," 2020.
- [6] S. Li, Y. Yao, J. Hu, G. Liu, X. Yao, and J. Hu, "An ensemble stacked convolutional neural network model for environmental event sound recognition," *Appl. Sci.*, vol. 8, no. 7, 2018.
- [7] C. Molnár et al., "Classification of dog barks: A machine learning approach," *Anim. Cogn.*, vol. 11, no. 3, pp. 389–400, 2008.
- [8] A. Maccagno, A. Mastropietro, U. Mazziotta, M. Scarpiniti, Y. C. Lee, and A. Uncini, "A CNN Approach for Audio Classification in Construction Sites," *Smart Innov. Syst. Technol.*, vol. 184, no. June, pp. 371–381, 2021.
- [9] B. Liu, Y. Liu, and K. Zhou, "Image Classification for Dogs and Cats," *IEEE Comput. Soc. Conf.*, vol. 1, no. 1, pp. 1–10, 2018.