**Jurnal Teknik Informatika (JUTIF)**
Vol. 3, No. 5, Oktober 2022, hlm. 1349-1353

DOI: https://doi.org/10.20884/1.jutif.2022.3.5.373
p-ISSN: 2723-3863
e-ISSN: 2723-3871

# *CLASSIFICATION OF CAT SOUNDS USING CONVOLUTIONAL NEURAL NETWORK (CNN) AND LONG SHORT-TERM MEMORY (LSTM) METHODS*

**Fadhilah Gusti Safinatunnajah[*1], Agi Prasetiadi[2], Merlinda Wibowo[3]**

[1,2,3]Program Studi Teknik Informatika, Fakultas Informatika, Institut Teknologi Telkom Purwokerto, Indonesia
Email: [1]fadhilahgusti1@gmail.com, [2]agi@ittelkom-pwt.ac.id, [3]merlinda@ittelkom-pwt.ac.id

(Naskah masuk: 10 Juni 2022, Revisi: 24 Juli 2022, diterbitkan: 24 Oktober 2022)

***Abstract***

*Cats become pets who are very close to humans, and they convey messages by producing identical sounds. Therefore, analysis of pet voices is important for a better relationship between cats and human. Animal communication through sound, especially in cats, depends on the situation or context in which the sound is made such as in a state of danger. Based on these problems, a classification method is needed to classify the similarity of characteristics in the resulting sound pattern. The classification methods used are Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) which can remember information for a long time and are used for a long time period. This study aimed to determine feelings or moods based on the sound produced into 4 categories: The Purr, The Meow, The Mating Call, and The Howl. The result of this study is that the best architectural model is to use 4 CNN convolution layers measuring 8-8-8-8 and 2 LSTM layers measuring 8-8. The precision value in this architecture is 0.68, the recall value is 1.00, the accurary value is 0.5625 and the f1-score value is 0.77. The small value of the confusion matrix is caused by the lack of dataset duration in the training process, resulting in underfitting.*

**Keywords**: *Cat, CNN, LSTM, Voice.*

# KLASIFIKASI SUARA KUCING MENGGUNAKAN METODE CONVOLUTIONAL NEURAL NETWORK (CNN) DAN LONG SHORT-TERM MEMORY (LSTM)

**Abstrak**

Kucing menjadi hewan peliharaan yang sangat dekat dengan manusia dan mereka menyampaikan pesan dengan menghasilkan suara yang identik oleh karena itu analisis suara hewan peliharaan menjadi penting agar terjalin hubungan antara kucing dengan manusia menjadi lebih baik. Komunikasi hewan melalui suara, khususnya kucing, tergantung kepada situasi atau konteks dimana suara itu dikeluarkan contohnya seperti dalam keadaan bahaya. Berdasarkan permasalah tersebut maka dibutuhkan sebuah metode klasifikasi yang dapat mengelompokkan kesamaan ciri pada pola suara yang dihasilkan. Metode klasifikasi yang digunakan yaitu *Convolutional Neural Network* (CNN) dan *Long Short-Term Memory* (LSTM) yang dapat mengingat informasi untuk waktu yang lama dan digunakan dalam jangka waktu yang panjang. Tujuan penelitian ini untuk mengetahui perasaan atau mood kucing berdasarkan suara yang dihasilkan menjadi 4 kategori yaitu *The Purr*, *The Meow*, *The Mating Call*, dan *The Howl*. Hasil dari penelitian ini adalah arsitektur model terbaik adalah dengan menggunakan 4 layer konvolusi CNN yang berukuran 8-8-8-8 dan 2 layer LSTM yang berukuran 8-8. Nilai *precision* pada arsitektur ini yaitu 0.68, nilai *recall* 1.00, nilai *accurary* 0.5625 dan nilai *f1-score* yaitu 0.77. Nilai confusion matrix yang kecil disebabkan oleh yang diakibatkan oleh kurangnya durasi dataset pada proses training sehingga terjadi *underfitting*.

**Kata kunci**: *CNN, Kucing, LSTM, Suara.*

## 1. PENDAHULUAN

Hubungan antara kucing dengan manusia sudah terjadi sejak tahun 8.000 Sebelum Masehi (SM) ketika manusia masih hidup berpindah. Pada awalnya, kucing berasal dari alam liar lalu perlahan-lahan mengalami proses domestikasi. Berdasarkan sejarah, usaha domestikasi kucing dimulai sekitar tahun 4.000 SM di Mesir. Saat itu kucing disukai karena kemampuannya dalam berburu tikus dan ular[1].

Sebagian besar hewan peliharaan khususnya kucing, menghabiskan seluruh waktunya di lingkungan manusia. Kucing rumahan adalah salah satu hewan peliharaan yang paling banyak dicintai di dunia dan populasinya sekitar 88,3 juta. Kini kucing menjadi hewan peliharaan yang sangat dekat dengan

manusia, dan mereka menyampaikan pesan dengan menghasilkan suara yang identik. 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 di lingkungan. Beberapa hewan memiliki kemampuan sensorik khusus, seperti penglihatan, pemandangan, perasaan, dan kesadaran akan perubahan alam dibandingkan dengan manusia. Suara binatang dapat bermanfaat bagi manusia dalam hal keamanan, prediksi bencana alam, dan interaksi intim jika kita mampu mengenalinya dengan baik [2] oleh karena itu analisis suara hewan peliharaan menjadi penting.

Cara kucing mengekspresikan suara mereka berbeda dengan manusia. Kita harus mempelajari perilaku mereka secara dekat sehingga kita dapat mempelajari komunikasi suara mereka dengan benar. Komunikasi hewan melalui suara, khususnya kucing, tergantung kepada situasi atau konteks dimana suara itu dikeluarkan. Kita harus mempelajari lingkungan sekitar ketika suara kucing dikeluarkan atau dibunyikan sehingga kita dapat mengenali pola tersebut. Kita dapat menganalisis dan menafsirkan alasan dari spesifik suara menghasilkan spesifik reaksi [3].

Dibutuhkan suatu metode untuk menafsirkan alasan atau arti dari suara kucing yang dihasilkan. Suara kucing dari kucing dapat diklasifikasikan menjadi 9 kategori yaitu *The Purr* (mendengkur), *The Trill* (getar), *The Meow* (meong), *The Howl* (melolong), *The Mating Call* (panggilan kawin), *The Growl* (menggeram), *The Hiss* (mendesis), *The Spit* (meludah), *The Snarl* (menggeram)-*Scream*, *Cry*, *Pain Shriek* , *The Chirp* (kicauan) dan *The Chatter* (obrolan). Setiap kategori memiliki arti yang berbeda-beda. Seperti, *The Purr* memiliki makna bahwa kucing melakukan *purr* tersebut ketika dalam konteks lapar, *stress*, dan kesakitan baik itu pada saat melahirkan atau ketika akan meninggal[3].

Penelitian mengenai klasifikasi suara sudah beberapa kali dilakukan seperti penelitian pada suara burung menggunakan CNN [4][5] dan juga suara kucing[2]. Metode CNN merupakan salah satu jenis neural network yang biasa digunakan dalam memproses data[6] yang terdiri dari beberapa layer[7]. Metode yang digunakan selain metode CNN pada klasifikasi suara yaitu metode LSTM seperti pada penelitian yang mendeteksi penyakit radang tenggorokan [8] dan penyakit jantung[9].

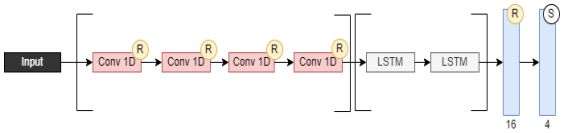
Berdasarkan masalah tersebut penelitian ini menggunakan metode *Convolutional Neural Network* (CNN) dan *Long Short Term Memory* (LSTM) untuk mengklasifikasikan suara kucing dengan 4 kategori yaitu *The Purr, The Meow*, *The Mating Call*, dan *The Howl*.

CNN terdiri dari beberapa layer [7], salah satunya yaitu convolution layer yang digunakan pada penelitian ini yang berfungsi sebagai filter untuk mencari ciri dari inputan[10]. LSTM (Long Short Term Memory) dipilih karena kemampuannya yang dapat belajar data dan mengingat informasi untuk waktu yang lama[11]

Penelitian ini menggunakan optimasi Adaptive Moment Estimation (ADAM) yang berfungsi untuk mengoptimalkan proses pembelajaran pada sistem[12]. Optimasi Adam dipilih karena kemampuannya yang dapat memperbarui nilai error dalam proses pelatihan[13].

## 2. METODE PENELITIAN

Metode penelitian ini menggunakan penggabungan 2 metode yaitu CNN dan LSTM dengan detail *layer* yang dapat dilihat pada gambar 2.1.



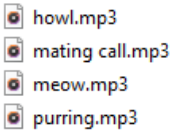

Gambar 1. Arsitektur Model Klasifikasi

Model klasifikasi dapat dilihat pada Gambar 1. Data yang sudah di *convert* menjadi file dengan format .mp3 kemudian di training menggunakan layer konvolusi CNN dengan aktivasi RelU untuk di ekstraksi fitur yang berguna untuk mengambil ciri dari inputan. Data lanjut ke layer LSTM yang berguna untuk mengenali bentuk sinyal suara yang merupakan data *sequencial*. LSTM memiliki lapisan memori untuk mengingat prediksi data yang dipelajari. Data kemudian masuk ke layer dense dengan layer pertama menggunakan aktivasi RelU dan layer kedua menggunakan aktivasi sigmoid untuk di kerucutkan filter hingga menjadi beberapa kategori saja sebagai penetuan klasifikasi yang mana pada penelitian ini suara kucing diklasifikasikan menjadi 4.

## 3. HASIL DAN PEMBAHASAN

### 3.1. Pengumpulan Bahan

Penelitian ini menggunakan data suara kucing yang didapatkan dari Youtube dengan total 4 kategori yaitu *The Purr*, *The Meow*, *The Howl*, dan *Mating Call*. Data suara kucing yang diunduh berupa file .mp4 yang kemudian disatukan menjadi 1 file sesuai kategori dan kemudian di convert menjadi .mp3.




Gambar 2. Dataset

Gambar 2 merupakan data yang sudah disatukan sehingga menghasilkan durasi yang berbeda-beda. Data yang akan digunakan pada model CNN dan

LSTM merupakan data suara kucing yang sudah disamakan durasinya yaitu 100 menit.

### 3.2. Membangun Model Klasifikasi

Arsitektur model yang digunakan terdiri dari CNN untuk bagian layer konvolusi yang terdiri dari 4 layer dan kemudian terdapat 2 layer LSTM. Aktivasi yang digunakan adalah RelU pada hidden layer, diakhiri dengan aktivasi Softmax pada layer pengambilan keputusan terakhir. Adapun optimalisasi menggunakan Adam.

Berikut merupakan skema dari arsitektur model klasifikasi yang dicoba:

1. Skema A

Skema A adalah skema untuk mencari Arsitektur CNN terbaik. Percobaan dilakukan sebanyak 6 kali dengan berbagai macam ukuran layer. Tabel 1 menunjukkan nilai akurasi dan validasi akurasi pada setiap model untuk mencari Arsitektur CNN terbaik.

Tabel 1 Skema Pencarian Arsitektur CNN

| No | Arsitektur | Akurasi | Val Akurasi |
|---|---|---|---|
| A.1 | CNN (4,4,4,4) + LSTM (8,8) | 0,8379 | 0.8708 |
| **A.2** | **CNN (8,8,8,8) + LSTM (8,8)** | **0,8515** | **0.8167** |
| A.3 | CNN (16,16,16,16) + LSTM (8,8) | 0,7678 | 0.7208 |
| A.4 | CNN (32,3,32,32,32) + LSTM (8,8) | 0.6527 | 0.6375 |
| A.5 | CNN (4,8,16,32) + LSTM (8,8) | 0,7803 | 0.5250 |
| A.6 | CNN (32,16,8,4) + LSTM (8,8) | 0,6789 | 0.7000 |

Tabel 1 menunjukkan bahwa arsitektur CNN terbaik adalah skema no A.2 yang memiliki 4 layer CNN berukuran 8,8,8,8 dan 2 layer LSTM berukuran 8,8 dengan nilai akurasi 0,8515 dan validasi akurasi 0,8167 pada proses training.

Penggunaan arsitektur CNN terbaik yaitu 8,8,8,8 kemudian dilanjutkan untuk mencari arsitektur LSTM terbaik pada skema B.

2. Skema B

Skema B adalah skema untuk mencari Arsitektur LSTM terbaik. Percobaan dilakukan sebanyak 4 kali dengan berbagai macam ukuran layer LSTM dan menggunakan 4 layer CNN yang sebelumnya sudah di cari yang terbaik pada skema A yaitu berukuran 8,8,8,8. Tabel 2 menunjukkan nilai akurasi dan validasi akurasi pada setiap model untuk mencari Arsitektur LSTM terbaik.

Tabel 2 Skema Pencarian Arsitektur LSTM

| No | Arsitektur | Akurasi | Val Akurasi |
|---|---|---|---|
| B.1 | CNN (8,8,8,8) + LSTM (4,4) | 0,7751 | 0.7792 |
| B.2 | CNN (8,8,8,8) + LSTM (4,8) | 0,6695 | 0.6667 |
| B.3 | CNN (8,8,8,8) + RNN (8,4) | 0,8138 | 0.8208 |
| **B.4** | **CNN (8,8,8,8) + RNN (8,8)** | **0,8515** | **0.8167** |

Tabel 2 menunjukkan bahwa arsitektur LSTM terbaik adalah skema no B.4 yang memiliki 4 layer CNN berukuran 8,8,8,8 dan 2 layer LSTM berukuran 8,8 dengan nilai akurasi 0,8515 dan validasi akurasi 0,8167 pada proses training.

Hasil dari skema A dan Skema B menunjukkan bahwa arsitektur terbaik untuk klasifikasi suara kucing ini menggunakan 4 layer CNN dengan ukuran 8,8,8,8 dan 2 layer LSTM dengan ukuran 8,8.

### 3.3. Fase Pengujian

Fase pengujian terhadap semua skema arsitektur menggunakan *confusion matrix* untuk mengevaluasi performa dari suatu model klasifikasi[14]. Berdasarkan confusion matrix tersebut dapat ditetapkan tolak ukur performa seperti Accuracy, Precision, Recall, Specificity, FMeasure, G-Mean dan yang lainnya[15].

Pengujian dilakukan dengan menggunakan dataset suara yang baru yang disebut data *testing* diluar data yang telah di *training*. Data testing terdiri dari 4 kategori yaitu *The Purr*, *The Mating Call*, *The Meow*, dan *The Howl* dengan masing-masing berdurasi 3 menit.

Hasil *confusion matrix* pada setiap model dari Skema A dan Skema B dapat dilihat pada Tabel 3 dibawah ini.

Tabel 3. Hasil Confusion Matrix

| No | Arsitektur | Precision | Recall | Accurary | F1-Score |
|---|---|---|---|---|---|
| A.1 | CNN (4,4,4,4) + RNN (8,8) | 0.13 | 0.50 | 0.28125 | 0.21 |
| **A.2** | **CNN (8,8,8,8) + RNN (8,8)** | **0.68** | **1.00** | **0.5625** | **0.77** |
| A.3 | CNN (16,16,16,16) + RNN (8,8) | 0.52 | 0.88 | 0.46875 | 0.64 |
| A.4 | CNN (32,3,32,32,32) + RNN (8,8) | 0.39 | 0.75 | 0.375 | 0.50 |
| A.5 | CNN (4,8,16,32) + RNN (8,8) | 0.11 | 0.38 | 0.1875 | 0.17 |
| A.6 | CNN (32,16,8,4) + RNN (8,8) | 0.29 | 0.62 | 0.3125 | 0.39 |
| B.1 | CNN (8,8,8,8) + RNN (4,4) | 0.62 | 0.88 | 0.4375 | 0.71 |
| B.2 | CNN (8,8,8,8) + RNN (4,8) | 0.23 | 0.56 | 0.28125 | 0.32 |
| B.3 | CNN (8,8,8,8) + RNN (8,4) | 0.67 | 0.94 | 0.53125 | 0.74 |
| **B.4** | **CNN (8,8,8,8) + RNN (8,8)** | **0.68** | **1.00** | **0.5625** | **0.77** |

Arsitektur terbaik berdasarkan Tabel 3 yaitu arsitektur dengan no skema A.2 atau B.4 dengan 4 Layer Konvolusi yang masing-masing layer berukuran 8,8,8, dan 8 dan 2 layer LSTM dengan masing-masing layer berukuran 8 dan 8, Aktivasi *RelU* pada *hidden layer*, diakhiri dengan aktivasi *Softmax* pada layer pengambilan keputusan terakhir dan optimalisasi menggunakan ADAM yang

berfungsi untuk mengoptimalkan sistem dengan cara memperbarui nilai *error* dalam proses *training*.

## 4. DISKUSI

Arsitektur skema A.2 merupakan yang terbaik pada penelitian ini dikarenakan jumlah data training yang digunakan sedikit sehingga dibutuhkan arsitektur yang sederhana dengan ukuran yang kecil.

Nilai *precision* pada skema arsitektur A.2 yaitu 0.68, nilai *recall* 1.00, nilai *accurary* 0.5625 dan nilai *f1-score* yaitu 0.77. Nilai *recall* 1.00 yang artinya bahwa jumlah data *testing* yang ditemukan kembali oleh sistem, sedangkan *accuracy* merupakan jumlah data yang diklasifikasikan secara benar oleh model arsitektur A.2 sehingga pada skema A.2 dapat disimpulkan terjadi *underfitting*, yakni ketika model *training* data yang dibuat tidak mewakili keseuruhan data yang digunakan sehingga model masih mempelajari hubungan antara variabel dalam data yang dalam penelitian ini dikarenakan kurangnya durasi dataset pada proses *training*.

Pada penelitian ini menyetujui hasil penelitian [16] dan [2] yang menyatakan bahwa Layer Convolusi pada metode CNN cocok untuk digunakan sebagai ekstraksi fitur dalam data dan hasil penelitian [9], [8], dan [17] menyatakan bahwa LSTM cocok digunakan untuk klasifikasi suara, terlebih jika digunakan ADAM untuk optimasi model, sehingga pada penelitian ini telah menguji bahwa semakin sedikit data training akan menyebabkan underfitting walaupun telah menggunakan CNN dan LSTM.

## 5. KESIMPULAN

Penelitian ini telah berhasil mengklasifikasikan suara kucing yang terdiri dari 4 kategori yaitu *The Purr*, *The Meow*, *The Howl*, dan *The Mating Call* menggunakan metode *Convolutional Neural Network* (CNN) dan *Long Short-Term Memory* (LSTM) dengan masing-masing kategori suara berdurasi 100 menit ketika di training.

Hasil kinerja dari metode Convolutional Neural Network (CNN) dan Long Short-Term Memory (LSTM) mendapatkan akurasi training 0.8515 atau 85.15% dengan arsitektur dengan 4 layer konvolusi CNN dengan ukuran 8-8-8-8 dan 2 layer LSTM dengan ukuran 8-8, Aktivasi *RelU* pada *hidden layer*, diakhiri dengan aktivasi *Softmax* pada layer pengambilan keputusan terakhir dan optimalisasi menggunakan ADAM. Nilai *precision* pada arsitektur ini yaitu 0.68, nilai *recall* 1.00, nilai *accurary* 0.5625 dan nilai *f1-score* yaitu 0.77. Nilai *confusion matrix* yang kecil disebabkan oleh yang diakibatkan oleh kurangnya durasi dataset pada proses *training* sehingga terjadi *underfitting*.

## DAFTAR PUSTAKA

[1] P. L. Bernstein, "The Human-Cat Relationship," pp. 47–89, 2007, doi: 10.1007/978-1-4020-3227-1_3.

[2] 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, doi: 10.3390/app8101949.

[3] S. Schötz, *The Secret Language of Cats: How to Understand Your Cat for a Better, Happier Relationship*. Hanover Square Press, 2018.

[4] C. Y. Koh, J. Y. Chang, C. L. Tai, D. Y. Huang, H. H. Hsieh, and Y. W. Liu, "Bird sound classification using convolutional neural networks," *CEUR Workshop Proc.*, vol. 2380, pp. 9–12, 2019.

[5] S. D. H. Permana, G. Saputra, B. Arifitama, Yaddarabullah, W. Caesarendra, and R. Rahim, "Classification of bird sounds as an early warning method of forest fires using Convolutional Neural Network (CNN) algorithm," *J. King Saud Univ. - Comput. Inf. Sci.*, no. xxxx, 2021, doi: 10.1016/j.jksuci.2021.04.013.

[6] Y. Luan and S. Lin, "Research on Text Classification Based on CNN and LSTM," *Proc. 2019 IEEE Int. Conf. Artif. Intell. Comput. Appl. ICAICA 2019*, pp. 352–355, 2019, doi: 10.1109/ICAICA.2019.8873454.

[7] C. Chen and F. Qi, "Single Image Super-Resolution Using Deep CNN with Dense Skip Connections and Inception-ResNet," *Proc. - 9th Int. Conf. Inf. Technol. Med. Educ. ITME 2018*, pp. 999–1003, 2018, doi: 10.1109/ITME.2018.00222.

[8] V. Guedes, A. Junior, J. Fernandes, F. Teixeira, and J. P. Teixeira, "Long short term memory on chronic laryngitis classification," *Procedia Comput. Sci.*, vol. 138, pp. 250–257, 2018, doi: 10.1016/j.procs.2018.10.036.

[9] Y. Chen, J. Lv, Y. Sun, and B. Jia, "Heart sound segmentation via Duration Long–Short Term Memory neural network," *Appl. Soft Comput. J.*, vol. 95, p. 106540, 2020, doi: 10.1016/j.asoc.2020.106540.

[10] Ridho Aji Pangestu, Basuki Rahmat, and Fetty Tri Anggraeny, "Implementasi Algoritma Cnn Untuk Klasifikasi Citra Lahan Dan Perhitungan Luas," *J. Inform. dan Sist. Inf.* , vol. 1, no. 1, pp. 166–174, 2020.

[11] X. H. Le, H. V. Ho, G. Lee, and S. Jung, "Application of Long Short-Term Memory (LSTM) neural network for flood forecasting," *Water (Switzerland)*, vol. 11, no. 7, 2019, doi: 10.3390/w11071387.

[12] P. O. Sgd, D. Irfan, R. Rosnelly, M. Wahyuni, J. T. Samudra, and A. Rangga, "MENGGUNAKAN CNN," vol. 4307, no. June, pp. 244–253, 2022.

[13] N. D. Miranda, L. Novamizanti, and S. Rizal, "Convolutional Neural Network Pada

Klasifikasi Sidik Jari Menggunakan Resnet-50," *J. Tek. Inform.*, vol. 1, no. 2, pp. 61–68, 2020, doi: 10.20884/1.jutif.2020.1.2.18.

[14] A. Amrin and H. Saiyar, "Aplikasi Diagnosa Penyakit Tuberculosis Menggunakan Algoritma Naive Bayes," *Jurikom*, vol. 5, no. 5, pp. 498–502, 2018.

[15] J. Yang, X. Huang, H. Wu, and X. Yang, "EEG-based emotion classification based on Bidirectional Long Short-Term Memory Network," *Procedia Comput. Sci.*, vol. 174, no. 2019, pp. 491–504, 2020, doi: 10.1016/j.procs.2020.06.117.

[16] C. Y. Koh, J. Y. Chang, C. L. Tai, D. Y. Huang, H. H. Hsieh, and Y. W. Liu, "Bird sound classification using convolutional neural networks," *CEUR Workshop Proc.*, vol. 2380, no. September, 2019.

[17] W. Zhang, J. Han, and S. Deng, "Abnormal heart sound detection using temporal quasi-periodic features and long short-term memory without segmentation," *Biomed. Signal Process. Control*, vol. 53, p. 101560, 2019, doi: 10.1016/j.bspc.2019.101560.