

# Klasifikasi Suara Paru Dengan Convolutional Neural Network (CNN)

Muhammad Hasbi Ashshiddieqy<sup>1</sup>, Jondri<sup>2</sup>, Achmad Rizal<sup>3</sup>

<sup>1,2,3</sup>Fakultas Informatika, Universitas Telkom, Bandung

<sup>1</sup>muhammadhasbia@students.telkomuniversity.ac.id, <sup>2</sup>jondri@telkomuniversity.ac.id,

<sup>3</sup>achmadrizal@telkomuniversity.ac.id

## Abstrak

Suara paru adalah suara yang disebabkan oleh pernapasan. Suara paru dapat menjadi salah satu pendeteksi kelainan pada pernapasan. Pendeteksian pada penelitian ini meliputi klasifikasi suara paru terhadap kelas normal, kelas *crackle*, kelas *wheeze*, dan kelas *crackle and wheeze*.

Spektrogram adalah representasi visual dari suara atau sinyal dengan bermacam frekuensi atau variabel lainnya. Spektrogram juga disebut sebagai short-time Fourier transform. Algoritma short-time Fourier transform (STFT) adalah algoritma pengembangan dari algoritma Transformasi Fourier. Kelebihan algoritma STFT dibandingkan algoritma Transformasi Fourier adalah STFT dapat menunjukkan lokasi terjadinya suatu frekuensi.

Penelitian ini menggunakan spektrogram karena algoritma pembelajaran mesin yang digunakan adalah convolutional neural network (CNN). CNN adalah algoritma yang telah tervalidasi dalam pengolahan citra medis. Oleh karena itu, penggunaan spektrogram akan membuat akurasi CNN menjadi lebih tinggi. Dalam penelitian ini digunakan tiga proses augmentasi, yaitu *shift up by a major third (four half-steps)*, *compress to be one and a half as fast*, *compress to be half as fast*. Proses augmentasi dilakukan untuk menghindari kondisi *overfitting* pada model pembelajaran mesin yang dilatih pada penelitian ini.

Setelah melakukan pelatihan pada model pembelajaran mesin CNN dengan menggunakan generalisasi berupa augmentasi dan *dropout layer*, maka didapatkan akurasi sebesar 84,80% terhadap data latih dan 78,09% terhadap data uji.

**Kata Kunci:** suara paru, spektrogram, CNN

## 1. Pendahuluan

### Latar Belakang

Pendeteksian penyakit pernapasan telah dilakukan dengan berbagai metode. Salah satu metode yang paling akurat adalah dengan melakukan klasifikasi terhadap suara pernapasan dengan melakukan pendeteksian terhadap kelainan. Klasifikasi kelainan pada suara paru dapat dilakukan dengan melakukan pemisahan antara suara pernapasan normal dan tidak normal (seperti *crackles* dan *wheezes*). Bunyi pernapasan memiliki peran penting dalam psikologi dan patologi dari paru-paru dan obstruksi saluran udara. Amplitudo dari suara paru dan kepadatan spektral dapat menunjukkan keadaan parenkim paru-paru, dimensi saluran udara, dan modifikasi patologi [1].

Deteksi kelainan pada suara paru biasa dilakukan dengan menggunakan stetoskop. Pendeteksian suara paru dengan stetoskop memiliki keterbatasan karena sangat bergantung pada individu yang melakukan pemeriksaan. Keterbatasan pada pemeriksaan dengan stetoskop dapat diatasi dengan penggunaan komputerisasi terhadap pembacaan otomatis terhadap pola akustik dari suara paru, sehingga klasifikasi suara paru dapat dilakukan dengan pendekatan objektif dan interpretasi yang telah terstandarisasi [2].

Rizal, dkk. telah melakukan penelitian terhadap klasifikasi suara paru dengan menggunakan fitur statistik orde pertama pada spektrogram. Pada penelitian tersebut, sinyal paru diubah menjadi spektrogram. Sinyal paru yang telah diubah menjadi spektrogram digunakan sebagai data untuk melakukan klasifikasi pada suara paru berdasarkan fitur yang terdapat pada spektrogram tersebut. Metode klasifikasi yang digunakan pada penelitian tersebut adalah K-Nearest Neighbor (KNN) [3].

Aykanat dkk. telah melakukan penelitian untuk klasifikasi suara paru dengan menggunakan dua ekstraksi fitur, yaitu ekstraksi fitur mel frequency cepstral coefficient (MFCC) dan pembuatan spektrogram dengan menggunakan short-time Fourier transform (STFT). Metode klasifikasi yang digunakan pada penelitian tersebut adalah Support Vector Machine (SVM) dan Convolutional Neural Network (CNN) [4].

Aridela dkk. telah melakukan penelitian terhadap suara paru dengan menggunakan data sinyal paru yang didekomposisi dengan wavelet Daubechies2 level 5. Metode klasifikasi yang digunakan pada penelitian tersebut adalah Support Vector Machine (SVM) dan klasifikasi jaringan syaraf tiruan Probabilistic Neural Network (PNN) [5].

Penelitian ini bertujuan untuk mendeteksi kelainan pernapasan pada suara paru. Deteksi kelainan pernapasan dilakukan dengan data berupa suara pernapasan yang telah dikonversi menjadi sebuah spektrogram dengan menggunakan algoritma short-time Fourier transform (STFT). Suara pernapasan yang telah dikonversi menjadi spektrogram akan dijadikan data latih untuk pembuatan model klasifikasi suara paru. Konversi suara paru dilakukan karena penelitian ini menggunakan algoritma convolutional neural network (CNN) dalam pembuatan model klasifikasi.

### Perumusan Masalah

Perumusan masalah yang dibahas dalam penelitian ini yaitu sebagai berikut.

- Apa ciri yang menandakan sebuah pernapasan memiliki kelainan?
- Bagaimana algoritma yang baik untuk klasifikasi suara paru?
- Apa perbedaan antara suara paru *normal*, *crackle*, *wheeze*, dan *crackle and wheeze*?

### Tujuan

Tujuan yang ingin dicapai dalam penelitian ini yaitu menganalisis suara paru yang menandakan sebuah pernapasan memiliki kelainan, mengimplementasikan algoritma yang baik dalam mengklasifikasi kelainan pada suara paru, melakukan klasifikasi terhadap suara paru normal, *crackle*, *wheeze*, dan *crackle and wheeze*.

### Organisasi Tulisan

Organisasi tulisan dalam penelitian ini yaitu sebagai berikut.

- Studi Terkait adalah bagian mengenai landasan teori yang digunakan pada penelitian.
- Sistem yang Dibangun adalah bagian mengenai prosedur pengerjaan penelitian.
- Evaluasi adalah bagian mengenai hasil dari penelitian beserta analisisnya.
- Kesimpulan adalah bagian yang berisi kesimpulan dari solusi terhadap tujuan penelitian
- Daftar Pustaka adalah bagian yang berisi sumber pustaka yang digunakan pada penelitian

## 2. Studi Terkait

### Suara Paru

Suara paru adalah suara yang dihasilkan oleh pernapasan yang terjadi pada tubuh. Suara paru terbagi menjadi tiga kategori : normal, abnormal, dan adventif [3]. Adventif merupakan kelainan pada paru yang terbagi menjadi dua jenis, suara adventif kontinu dan suara adventif tidak kontinu [6].

### Spektrogram

Spektrogram adalah representasi visual dari suara atau sinyal dengan bermacam frekuensi atau variabel lainnya [4]. Teknik spektrogram dilakukan untuk mengubah sinyal dari time domain menjadi time-frequency domain.

Penggunaan spektrogram adalah untuk mengatasi kelemahan yang terdapat pada transformasi Fourier. Kelemahan dari transformasi Fourier adalah tidak dapat menentukan lokasi terjadinya frekuensi [7]. Spektrogram disebut

sebagai short-time Fourier transform (STFT) karena spektrogram melakukan transformasi Fourier pada interval ketika sinyal dianggap diam. Short-time Fourier transform dapat dirumuskan sebagai berikut:

$$X(t, f) = \int_{-\infty}^{\infty} x(\tau)w(t - \tau)e^{-j2\pi mf\tau} d\tau$$

Dimana  $w(t-\tau)$  adalah fungsi jendela yang digunakan untuk melakukan pemotongan dan filter pada sinyal. Panjang dari fungsi jendela akan berpengaruh pada resolusi waktu dan resolusi frekuensi pada spektrogram [3].

### Convolutional Neural Network (CNN)

Convolutional neural network (CNN) adalah salah satu kelas deep feed-forward artificial neural networks yang terinspirasi oleh proses-proses biologi dengan pola konektivitas antar neuron menyerupai organisasi visual cortex pada binatang [8]. CNN adalah representasi gambar dan teknik yang telah tervalidasi untuk melakukan analisis dan aplikasi terhadap citra medis [9].

Convolutional neural network memiliki enam lapisan pembangun, yaitu : convolutional layer, pooling layer, normalization layer, Rectified Linear Units layer, fully connected layer, dan loss layer [8].

Convolutional layer merupakan lapisan inti pada CNN. Pada lapis ini seluruh bidang reseptif akan ditelusuri oleh filter secara tumpang tindih parsial. Penelusuran filter secara tumpang tindih parsial menyebabkan setiap neuron akan berbagi bobot koneksi (weight sharing) [8].

Pooling layer merupakan lapisan untuk meringkas output dari setiap grup neuron pada lokasi kernel yang sama [10]. Pooling layer dilakukan untuk menjaga ukuran data ketika melakukan konvolusi dengan cara melakukan downsampling. Dengan lapis ini, representasi data menjadi lebih kecil, mudah dikelola, dan mudah dalam mengendalikan overfitting [8].

Normalization layer digunakan untuk melakukan normalisasi pada citra masukan. Normalisasi pada citra masukan bertujuan untuk mengatasi perbedaan rentang nilai yang signifikan

Rectified Linear Units (ReLU) layer berfungsi untuk meningkatkan sifat nonlinearitas fungsi keputusan dan jaringan secara keseluruhan tanpa mempengaruhi bidang reseptif pada lapis konvolusi [8].

Fully connected layer merupakan lapis yang memiliki kesamaan dengan Multi Layer Perceptron (MLP). Pada lapis ini dilakukan perkalian matriks yang diikuti dengan bias offset [8].

## 3. Sistem yang dibangun

### Praproses

Data yang diberikan berupa data sekunder yang terbagi menjadi tiga jenis, yaitu file audio yang berisikan rekaman suara paru, file teks yang berisi label target persiklus pernafasan, dan file teks yang berisi tempat dilakukannya perekaman suara paru. Pada kasus klasifikasi suara paru data terbagi menjadi empat kelas berbeda yaitu kelas normal sejumlah 1251 data (52.81%), kelas *crackle* sejumlah 642 data (27.1%), kelas *wheeze* sejumlah 304 data (12.83%), dan kelas *crackle and wheeze* sejumlah 172 data (7.26%). Praproses yang pertama dilakukan adalah menggabungkan seluruh file teks yang berisi label target pernafasan dan melakukan encode terhadap label. Proses ini dilakukan untuk mempermudah pemotongan file audio menjadi persiklus pernafasan. Pemotongan terhadap file audio dilakukan karena pelabelan dilakukan terhadap setiap siklus pernafasan yang terjadi.

### Augmentasi

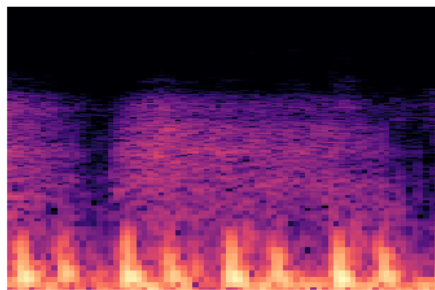
Augmentasi merupakan salah satu proses yang bertujuan untuk mencegah terjadinya *overfitting*. Data audio yang telah dipotong menjadi persiklus pernafasan, kemudian diaugmentasi. Proses augmentasi bertujuan untuk menambah varian data pada saat proses pelatihan model. Melakukan augmentasi adalah salah satu langkah regularisasi yang dilakukan untuk mencegah terjadinya kondisi *overfitting*. Dengan dilakukannya augmentasi pada data, maka akan mencegah terjadinya pembuatan model yang terlalu spesifik. Proses augmentasi yang dilakukan pada data audio yang telah melalui praproses adalah menambah kecepatan pernafasan menjadi satu setengah kali lipat, mengurangi kecepatan pernafasan menjadi setengah kali lipat, dan dengan melakukan *pitch shifting*.

### Short-Time Fourier Transform

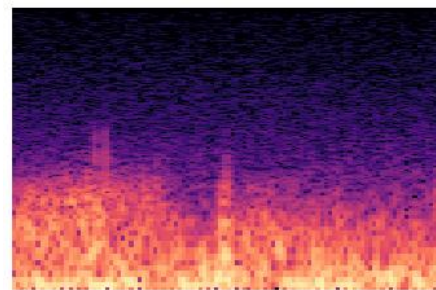
Short-Time Fourier Transform merupakan algoritma transformasi yang digunakan untuk mendekomposisi suatu sinyal kompleks menjadi sinyal sinus yang lebih sederhana. STFT akan merubah sinyal menjadi time-frequency domain. Penggunaan STFT dilakukan karena proses klasifikasi suara paru akan dilakukan dengan menggunakan spektrogram sebagai data untuk melakukan pelatihan model. Pada beberapa riset perbandingan dalam pembuatan sebuah spektrogram, terbukti bahwa melakukan pembuatan spektrogram dengan STFT akan memberikan kinerja terbaik terhadap proses klasifikasi.

### Spektrogram

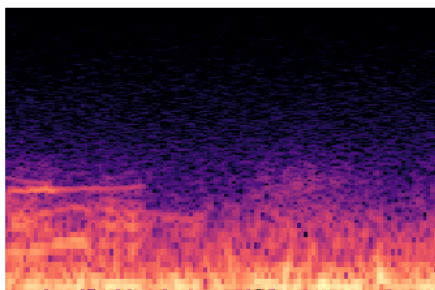
Setelah melakukan transformasi sinyal dengan menggunakan STFT, hasil transformasi tersebut kemudian ditampung menjadi spektrogram. Spektrogram digunakan untuk merubah representasi sinyal yang pada awalnya adalah 1D, menjadi representasi sinyal dengan 2D. Perubahan representasi sinyal dari 1D menjadi 2D dilakukan karena algoritma pembelajaran *Convolutional Neural Network* (CNN) akan memiliki kinerja yang lebih baik pada klasifikasi citra 2D.



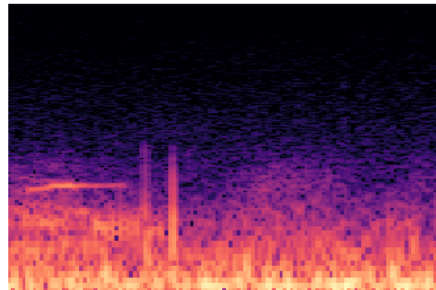
Normal



Crackle



Wheeze



Crackle and Wheeze

### Model *Convolutional Neural Network* dan Hyperparameter

Pembuatan model CNN dilakukan dengan penentuan arsitektur yang akan digunakan. Penentuan arsitektur dilakukan dengan memperhatikan faktor-faktor yang akan memungkinkan terjadinya *overfitting*. Dalam pencegahan terjadinya *overfitting* maka diterapkan proses regularisasi pada model dengan menggunakan *dropout layer*. Setelah terbentuknya arsitektur CNN yang akan digunakan, maka dilakukanlah penentuan *hyperparameter* melalui beberapa proses validasi terhadap kinerja model CNN dan *hyperparameter* yang ditentukan.

### KFold Cross-Validation

Suatu model tidak dapat dikatakan baik apabila belum stabil. Proses validasi dengan *kfold cross-validation* dilakukan sebagai penentu bahwa suatu model sudah dapat dikatakan stabil. Validasi dilakukan dengan menggunakan *cross-validation* karena data yang digunakan pada pelatihan model sejumlah 2369 data dan jumlah tersebut termasuk ke dalam kategori *small data set*. Proses *kfold cross-validation* yang dilakukan adalah dengan membagi himpunan data menjadi sepuluh bagian (*fold*) dengan bertingkat (10 *fold stratified cross-validation*).

|              |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
|--------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Fold         | 1     |       |       | 2     |       |       | 3     |       |       | 4     |       |       | 5     |       |       |
| Accuracy (%) | 63.87 | 67.78 | 62.37 | 66.78 | 68.33 | 67.82 | 62.49 | 70.56 | 60.89 | 65.43 | 63.89 | 65.35 | 64.18 | 67.22 | 63.86 |
| Fold         | 6     |       |       | 7     |       |       | 8     |       |       | 9     |       |       | 10    |       |       |
| Accuracy (%) | 64.75 | 65.56 | 63.36 | 64.83 | 69.44 | 64.35 | 64.84 | 64.44 | 65.34 | 66.21 | 71.11 | 63.36 | 63.34 | 60.56 | 60.39 |

|  |       |            |      |            |
|--|-------|------------|------|------------|
|  | Train | Validation | Test | epoch = 15 |
|--|-------|------------|------|------------|

Berdasarkan tabel tersebut dapat disimpulkan bahwa model CNN yang dibuat dapat dinyatakan sebagai model yang stabil. Kesimpulan tersebut diambil berdasarkan tidak terjadinya *overfitting* pada setiap pergantian *fold*. Selain itu, model tersebut juga memiliki akurasi yang cukup tinggi pada setiap pergantian *fold*.

**Pelatihan Model**

Pelatihan model dilakukan apabila suatu model sudah memiliki kinerja yang baik dan stabil. Model yang stabil adalah model yang tidak memiliki kemungkinan *overfitting* ketika melalui proses validasi. Pelatihan model dilakukan dengan menggunakan 60 *epoch* dan dengan pembagian 66.27% data latih, 18.7% data validasi, dan 15.07 data tes.

**Evaluasi Model**

Proses evaluasi model adalah proses pengujian yang dilakukan untuk mengetahui kinerja suatu model apabila diaplikasikan pada kasus yang belum pernah digunakan dalam proses pelatihan (*unseen data*). Proses evaluasi sangat penting untuk mengetahui bahwa suatu model memiliki kinerja yang baik tidak hanya pada suatu himpunan data tertentu.

**Confusion Matrix**

*Confusion matrix* adalah metrik evaluasi yang umum digunakan pada kasus klasifikasi. Pemilihan *confusion matrix* sebagai metrik evaluasi dilakukan untuk mengetahui kinerja model pada proses klasifikasi untuk setiap kelas yang telah ditentukan pada kasus klasifikasi suara paru.

**4. Evaluasi**

**Hasil Pengujian**

|              |                |       |       |                |       |       |                |       |       |                                      |  |  |  |  |  |
|--------------|----------------|-------|-------|----------------|-------|-------|----------------|-------|-------|--------------------------------------|--|--|--|--|--|
| Architecture | Architecture 1 |       |       | Architecture 2 |       |       | Architecture 3 |       |       | Accuracy was gotten from best weight |  |  |  |  |  |
| Accuracy (%) | 84.8           | 77.47 | 78.09 | 86             | 76.75 | 79.21 | 52.8           | 52.82 | 52.81 |                                      |  |  |  |  |  |

|               |       |       |       |       |       |       |         |      |       |        |       |       |         |       |       |
|---------------|-------|-------|-------|-------|-------|-------|---------|------|-------|--------|-------|-------|---------|-------|-------|
| Learning Rate | 0.002 |       |       | 0.001 |       |       | 0.00075 |      |       | 0.0005 |       |       | 0.00025 |       |       |
| Accuracy (%)  | 67.84 | 69.89 | 68.82 | 70.71 | 71.95 | 68.54 | 89.85   | 75.4 | 79.21 | 84.8   | 77.47 | 78.09 | 99.27   | 76.55 | 78.37 |

|  |       |            |      |
|--|-------|------------|------|
|  | Train | Validation | Test |
|--|-------|------------|------|

Pada tabel pertama dilakukan perbandingan terhadap tiga arsitektur CNN. Arsitektur 1 adalah arsitektur yang dibuat dengan urutan *convolution layer* -> *pooling layer* sejumlah empat lapisan pada lapisan konvolusi dan *fully connected layer* sejumlah tiga lapisan pada lapisan klasifikasi. Arsitektur 2 adalah arsitektur yang dibuat dengan urutan *convolution layer* -> *pooling layer* sejumlah enam lapisan pada lapisan konvolusi dan *fully connected layer* sejumlah empat lapisan pada lapisan klasifikasi. Arsitektur 3 adalah arsitektur yang dibuat dengan urutan *convolution layer* -> *convolution layer* -> *pooling layer* sejumlah enam lapisan pada lapisan konvolusi dan *fully connected layer* sejumlah empat lapisan pada lapisan klasifikasi.

Berdasarkan hasil pengujian terhadap tiga arsitektur model mesin pembelajaran, dapat disimpulkan bahwa arsitektur 1 dan arsitektur 2 memiliki kinerja yang baik dalam melakukan klasifikasi terhadap suara paru. Maka kita dapat memilih di antara kedua arsitektur tersebut, namun yang dapat dipertimbangkan adalah perbedaan kinerja model pada data latih, data validasi, dan data uji. Sehingga dapat diambil kesimpulan bahwa arsitektur 1 memiliki kinerja yang lebih baik. Setelah melakukan pemilihan arsitektur, dilakukan perbandingan terhadap nilai *learning rate* dengan menggunakan arsitektur 1. Dari hasil perbandingan kinerja model terhadap lima nilai *learning rate*, maka dapat disimpulkan bahwa nilai *learning rate* terbaik terdapat pada saat *learning rate* bernilai 0.0005.

### Analisis Hasil Pengujian

Dari beberapa arsitektur yang digunakan sebagai pengujian kinerja model klasifikasi dapat disimpulkan bahwa semakin dalam model akan mempercepat proses pembelajaran, tetapi hal tersebut juga dapat menyebabkan kurang stabilnya model karena memperbesar kemungkinan terjadinya *overfitting*.

Berdasarkan hasil dari pengujian terhadap pencarian nilai *learning rate* dengan hasil kinerja terbaik dapat disimpulkan bahwa *learning rate* dengan nilai yang tinggi kurang efisien ketika digunakan pada kasus citra dengan pola yang kompleks, tetapi nilai *learning rate* yang terlalu rendah memungkinkan terjadinya pembacaan *noise* yang kemudian dianggap sebagai fitur dari citra.

Selain itu, keseimbangan jumlah data juga termasuk sebagai faktor penting untuk membuat model pembelajaran mesin dengan kinerja yang baik. Pada kasus klasifikasi suara paru data terbagi menjadi empat kelas berbeda yaitu kelas normal sejumlah 1251 data (52.81%), kelas *crackle* sejumlah 642 data (27.1%), kelas *wheeze* sejumlah 304 data (12.83%), dan kelas *crackle and wheeze* sejumlah 172 data (7.26%). Dari pembagian data tersebut dapat disimpulkan bahwa salah satu faktor yang membuat kinerja model menjadi tidak optimum adalah tidak seimbang jumlah data pada setiap kelas.

## 5. Kesimpulan

Berdasarkan pengujian yang telah dilakukan maka dapat disimpulkan :

1. Analisis kelainan suara paru dapat dilakukan dengan lebih baik apabila suara paru telah ditransformasi dan direpresentasikan dengan spektrogram. Maka pada penelitian ini, suara paru berhasil dianalisis melalui spektrogram yang telah dibuat.
2. Implementasi yang baik dalam klasifikasi suara paru adalah dengan menggunakan spektrogram dan algoritma pembelajaran *convolutional neural network* (CNN). Penggunaan spektrogram mempermudah algoritma pembelajaran CNN untuk melakukan ekstraksi fitur pada gambar dan melakukan klasifikasi berdasarkan fitur tersebut.
3. Klasifikasi terhadap suara paru normal, *crackle*, *wheeze*, dan *crackle and wheeze* dapat dilakukan setelah mengimplementasikan algoritma yang baik dalam penelitian ini. Klasifikasi terhadap suara paru menggunakan CNN memiliki hasil kinerja yang baik.

### Daftar Pustaka

- [1] S. Reichert, R. Gass, C. Brandt, and E. Andr s, "Analysis of Respiratory Sounds: State of the Art," *Clin. Med. Circ. Respirat. Pulm. Med.*, vol. 2, p. CCRPM.S530, 2008.
- [2] C. Pinho, A. Oliveira, C. J cume, J. Rodrigues, and A. Marques, "Automatic Crackle Detection Algorithm Based on Fractal Dimension and Box Filtering," *Procedia Comput. Sci.*, vol. 64, pp. 705–712, 2015.
- [3] A. Rizal, R. Hidayat, and H. A. Nugroho, "Lung Sounds Classification using Spectrogram 's First Order Statistics Features," 2016.
- [4] M. Aykanat,  . Kili , B. Kurt, and S. Saryal, "Classification of lung sounds using convolutional neural networks," *EURASIP J. Image Video Process.*, vol. 2017, no. 1, p. 65, Dec. 2017.
- [5] C. Aridela *et al.*, "Perbandingan Suara Paru Normal Dan Abnormal Menggunakan Probabilistic Neural Network Dan Support Vector Machine Classification Comparison of Normal and Abnormal Lung Sounds," vol. 4, no. 1, pp. 165–174, 2017.

- [6] R. Palaniappan, K. Sundaraj, and N. U. Ahamed, "Machine learning in lung sound analysis: A systematic review," *Biocybern. Biomed. Eng.*, vol. 33, no. 3, pp. 129–135, 2013.
- [7] B. Boashash, "Time-Frequency Signal Analysis and Processing: A Comprehensive Reference," Jan. 2003.
- [8] Suyanto, *Machine Learning Tingkat Dasar dan Lanjut*. 2018.
- [9] L. Lu, Y. Zheng, G. Carneiro, and L. Yang, *Advances in Computer Vision and Pattern Recognition Deep Learning and Convolutional Neural Networks for Medical Image Computing*. .
- [10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks."