

**REKOGNISI BUNYI KATA BERBAHASA INDONESIA  
MENGUNAKAN TENSORFLOW DAN METODE  
CONVOLUTIONAL NEURAL NETWORK**

**SKRIPSI**

Untuk Memenuhi Sebagian Persyaratan Mencapai Derajat Sarjana S-1

Program Studi Teknik Informatika



**Disusun oleh:**

**Fayyadh Aunilbarr**

**18106050043**

**STATE ISLAMIC UNIVERSITY  
SUNAN KALIJAGA  
YOGYAKARTA  
PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA  
YOGYAKARTA**

**2022**

# HALAMAN PENGESAHAN



KEMENTERIAN AGAMA  
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA  
FAKULTAS SAINS DAN TEKNOLOGI  
Jl. Marsda Adisucipto Telp. (0274) 540971 Fax. (0274) 519739 Yogyakarta 55281

## PENGESAHAN TUGAS AKHIR

Nomor : B-861/Un.02/DST/PP.00.9/05/2022

Tugas Akhir dengan judul : REKOGNISI BUNYI KATA BERBAHASA INDONESIA MENGGUNAKAN  
TENSORFLOW DAN METODE CONVOLUTIONAL NEURAL NETWORK

yang dipersiapkan dan disusun oleh:

Nama : FAYYADH AUNILBARR  
Nomor Induk Mahasiswa : 18106050043  
Telah diujikan pada : Jumat, 08 April 2022  
Nilai ujian Tugas Akhir : A-

dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta

### TIM UJIAN TUGAS AKHIR



Ketua Sidang  
Mandahadi Kusuma, M.Eng.  
SIGNED

Valid ID: 6252566098292



Penguji I  
Ir. Maria Ulfah Siregar, S.Kom., MIT., Ph.D.  
SIGNED

Valid ID: 627a4ae5a7e9c



Penguji II  
Dr. Ir. Shofwatul 'Uyun, S.T., M.Kom.  
SIGNED

Valid ID: 6279e2c733013



Yogyakarta, 08 April 2022  
UIN Sunan Kalijaga  
Dekan Fakultas Sains dan Teknologi  
Dr. Dra. Hj. Khurul Wardati, M.Si.  
SIGNED

Valid ID: 627c7fb1eece4a

## SURAT PERSETUJUAN SKRIPSI/TUGAS AKHIR



Universitas Islam Negeri Sunan Kalijaga



FM-UINSK-BM-05-03/R0

### SURAT PERSETUJUAN SKRIPSI/TUGAS AKHIR

Hal : Persetujuan Skripsi

Lamp :

Kepada

Yth. Dekan Fakultas Sains dan Teknologi

UIN Sunan Kalijaga Yogyakarta

di Yogyakarta

*Assalamu'alaikum wr. wb.*

Setelah membaca, meneliti, memberikan petunjuk dan mengoreksi serta mengadakan perbaikan seperlunya, maka kami selaku pembimbing berpendapat bahwa skripsi Saudara:

Nama : Fayyadh Aunilbarr

NIM : 18106050043

Judul Skripsi : Rekognisi Bunyi Kata Berbahasa Indonesia Menggunakan Tensorflow dan metode Convolutional Neural Network,

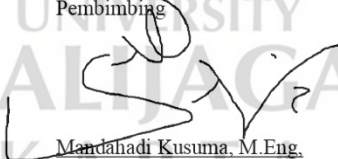
sudah dapat diajukan kembali kepada Program Studi Teknik Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata Satu dalam Program Studi Teknik Informatika

Dengan ini kami mengharap agar skripsi/tugas akhir Saudara tersebut di atas dapat segera dimunaqsyahkan. Atas perhatiannya kami ucapkan terima kasih.

*Wassalamu'alaikum wr. wb.*

Yogyakarta, 24 Maret 2022

Pembimbing

  
Mandahadi Kusuma, M.Eng.  
NIP. 19841115 201903 1 003

## PERNYATAAN KEASLIAN SKRIPSI

### PERNYATAAN KEASLIAN SKRIPSI

Saya yang bertanda tangan di bawah ini:

Nama : Fayyadh Aunilbar  
NIM : 18106050043  
Program Studi : Teknik Informatika  
Fakultas : Sains dan Teknologi

Menyatakan bahwa skripsi saya yang berjudul **“Rekognisi Bunyi Kata Berbahasa Indonesia Menggunakan Tensorflow Dan Metode Convolutional Neural Network”** merupakan hasil penelitian saya sendiri, tidak terdapat pada karya yang pernah diajukan untuk memperoleh gelar kesarjana di suatu perguruan tinggi, dan bukan plagiasi karya orang lain kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 24 Maret 2022



Fayyadh Aunilbar  
NIM. 18106050043

STATE ISLAMIC UNIVERSITY  
SUNAN KALIJAGA  
YOGYAKARTA

## KATA PENGANTAR

Dengan memanjatkan puji syukur kepada Allah SWT atas limpahan rahmat, hidayah, serta karuniaNya kepada penulis sehingga penulis dapat menyelesaikan tugas akhir sebagai persyaratan meraih gelas Sarjana Strata 1 dengan judul :

**“REKOGNISI BUNYI KATA BERBAHASA INDONESIA  
MENGUNAKAN TENSORFLOW DAN METODE  
CONVOLUTIONAL NEURAL NETWORK”**

Di dalam penulisan ini disajikan pokok-pokok bahasan yang meliputi latar belakang, metode penelitian, serta hasil dan pembahasan yang telah didapat pada proses penelitian. Kesimpulan penelitian didapatkan berdasarkan hasil pada proses penelitian ini.

Penulis menyadari bahwa penulisan tugas akhir ini masih ada kekurangan di dalamnya. Untuk itu, penulis mengharapkan kritik serta saran yang membangun untuk menyempurnakan penulisan tugas akhir ini. Sehingga, dapat bermanfaat bagi berbagai pihak dan juga bagi perkembangan ilmu pengetahuan terutama di bidang informatika.

Yogyakarta, 24 Maret 2022



Fayyadh Aunilbarr

## HALAMAN PERSEMBAHAN

*Skripsi ini saya persembahkan untuk almamater dan keluarga tercinta yang tidak henti memberikan energi positif kepada penulis untuk terus tumbuh menjadi pribadi yang lebih baik.*

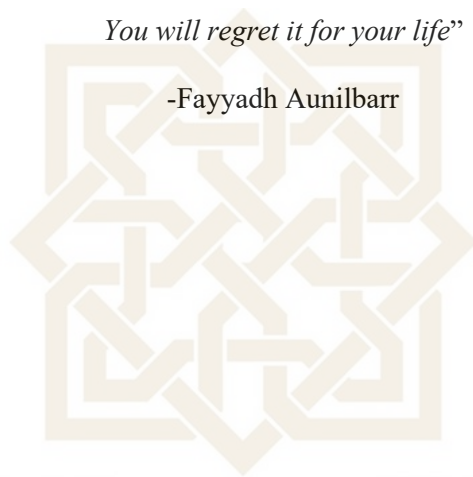


## HALAMAN MOTTO

*“If you never try even once,*

*You will regret it for your life”*

-Fayyadh Aunilbarr



STATE ISLAMIC UNIVERSITY  
SUNAN KALIJAGA  
YOGYAKARTA

## DAFTAR ISI

HALAMAN PENGESAHAN .....	ii
SURAT PERSETUJUAN SKRIPSI/TUGAS AKHIR.....	iii
PERNYATAAN KEASLIAN SKRIPSI .....	iv
KATA PENGANTAR .....	v
HALAMAN PERSEMBAHAN.....	vi
HALAMAN MOTTO.....	vii
DAFTAR ISI .....	viii
DAFTAR GAMBAR .....	x
DAFTAR TABEL.....	xi
INTISARI.....	xii
ABSTRACT .....	xiii
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah .....	2
1.3 Tujuan Penelitian.....	3
1.4 Batasan Masalah .....	3
1.5 Manfaat Penelitian.....	3
1.6 Kebaruan Penelitian .....	3
BAB II STUDI PUSTAKA.....	5
BAB III LANDASAN TEORI .....	8
3.1 Pengenalan suara .....	8
3.2 Audio.....	8
3.3 Convolutional Neural Networks.....	9
3.4 Convolutional Layer .....	9
3.5 Max Pooling Layer.....	10
3.6 Dropout .....	11
3.7 Confusion Matrix.....	11
BAB IV METODOLOGI PENELITIAN.....	12
4.1 Alat dan Bahan.....	12
4.2 Alur Penelitian .....	13
4.3 Metode Penelitian .....	15



<b>BAB V HASIL DAN PEMBAHASAN .....</b>	<b>16</b>
<b>5.1 Pengumpulan Data .....</b>	<b>16</b>
<b>5.2 Preprocess data .....</b>	<b>19</b>
<b>5.3 Implementasi CNN .....</b>	<b>21</b>
<b>5.4 Analisa dan Evaluasi .....</b>	<b>24</b>
<b>BAB VI PENUTUP.....</b>	<b>30</b>
<b>6.1 Kesimpulan.....</b>	<b>30</b>
<b>6.2 Saran .....</b>	<b>30</b>
<b>DAFTAR PUSTAKA .....</b>	<b>32</b>
<b>LAMPIRAN KODE PROGRAM .....</b>	<b>35</b>



## DAFTAR GAMBAR

Gambar 4.1 Alur Penelitian.....	13
Gambar 5.1.1 Contoh data suara dalam bentuk format .....	18
Gambar 5.2.1 Kode program preprocess data .....	19
Gambar 5.2.2 Kode program mengubah waveform menjadi spectogram .....	20
Gambar 5.2.3 Contoh hasil konversi spectogram dari waveform ....	20
Gambar 5.3.1 Kode program membuat model CNN .....	21
Gambar 5.3.2 Kode program compile model.....	22
Gambar 5.3.3 Kode program pelatihan model .....	23
Gambar 5.4.1 Hasil Analisa loss dan accuracy dengan rasio 60:20:20 .....	24
Gambar 5.4.2 Hasil Analisa loss dan accuracy dengan rasio 70:15:15 .....	25
Gambar 5.4.3 Hasil Analisa loss dan accuracy dengan rasio 70:10:10 .....	26
Gambar 5.4.4 Hasil confusion matrix dengan rasio 60:20:20 .....	27
Gambar 5.4.5 Hasil confusion matrix dengan rasio 70:15:15 .....	27
Gambar 5.4.6 Hasil confusion matrix dengan rasio 70:10:10 .....	27

## DAFTAR TABEL

Tabel 5.1.1 Kriteria pengisian suara .....	16
Tabel 5.1.2 Pengisian suara oleh relawan masing-masing .....	17
Tabel 5.4.1 Hasil akurasi antara 3 rasio yang berbeda .....	26
Tabel 5.4.2 Confusion matrix rasio 60:20:20.....	28
Tabel 5.4.3 Confusion matrix rasio 70:15:15.....	28
Tabel 5.4.4 Confusion matrix rasio 80:10:10.....	29



## INTISARI

Pengenalan (voice recognition) yang memungkinkan komputer mampu memproses suara atau sinyal *audio* menjadi teks. Ini dilakukan pada komputer maupun perangkat seluler pintar yang mampu mengenali suara tanpa menyentuh layar seperti *Siri* yang tersedia pada perangkat iOS, Google Now pada perangkat Android, atau Google Docs *Voice Typing* untuk menulis dokumen hanya dengan ucapan dari pengguna. Pusat Layanan Difabel merupakan Lembaga yang melayani mahasiswa dengan disabilitas, termasuk tuli atau tunarungu, di lingkungan Universitas Islam Negeri Sunan Kalijaga.

Tujuan dari penelitian ini adalah melakukan rekognisi bunyi kata berbahasa Indonesia dengan menggunakan *Convolutional Neural Network*. CNN merupakan salah satu metode dalam *Deep Learning* yang merupakan salah satu bagian dari *Artificial Intelligence* atau Kecerdasan Buatan. CNN adalah salah satu metode populer digunakan untuk klasifikasi gambar. Dataset terdiri dari 4 kata yang berbeda, masing-masing memiliki 50 data suara. Sumber data suara merupakan relawan dari PLD, baik laki-laki maupun perempuan, di antara usia 19 sampai dengan 23 tahun. Penulis menggunakan cara *preprocessing* data dengan konversi audio menjadi *spectrogram* yang dapat digunakan sebagai *input* gambar untuk pelatihan model dengan metode CNN yang hanya bisa menerima *input* gambar. *Image resizing* dan *spectrogram* dapat mengurangi kualitas dari seutuhnya data audio karena pelatihan model dapat dilakukan lebih cepat jika menggunakan ukuran yang lebih kecil dan sama, dan *spectrogram* hanya menampilkan frekuensi dari waktu ke waktu tanpa informasi amplitudo dari waktu ke waktu.

Meskipun kekurangan pada *preprocessing* data, hasil penelitian ini menunjukkan metode CNN dengan rasio dari data latih, data validasi, data uji sebesar 60:20:20 lebih baik dari rasio 70:15:15 dan 80:10:10 dalam akurasi dari data uji. Penelitian ini menggunakan 50 *epochs* untuk melakukan pelatihan model CNN tetap membuahkan hasil tes akurasi sebesar 80% keatas. Untuk hasil tes akurasi yang paling besar adalah menggunakan rasio 60:20:20 dengan hasil tes akurasi sebesar 90%. Penulis menyarankan penelitian selanjutnya dapat menunjukkan perbandingan metode lain seperti ANN atau RNN dengan menggunakan dataset berbahasa Indonesia dan menggunakan *denoise* pada tahap *preprocessing*.

## ABSTRACT

Voice recognition that allows computers to process voice or audio signals into text or do other things. This is done on computers or smart mobile devices that are able to recognize voice without touching layers such as Siri available on iOS devices, Google Now on Android devices, or Google Docs Voice Typing to write documents with just the author's speech. Pusat Layanan Difabel is an institution that serves students with disabilities, including the Deaf or Hard of Hearing, within the State Islamic University of Sunan Kalijaga.

The purpose of this research is to recognize the sounds of Indonesian words using the Convolutional Neural Network. CNN is one of the methods in Deep Learning which is a part of Artificial Intelligence. CNN is a popular method used for image classification. The dataset consists of 4 different words, each containing 50 voice data. Voice data sources are volunteers from PLD, both male and female, between the ages of 19 to 23 years old. The author will use preprocessing data by converting audio into a spectrogram that can be used as image input for training model with the CNN method which can only accept image input. Image resizing and spectrogram can reduce the quality of the original audio data because training model can performs much faster if using smaller and equal sizes, and spectrograms only display frequency over time without amplitude information over time.

Despite the shortcomings in data preprocessing, the result of this study shows that the CNN method with a ratio of training data, validation data, and test data of 60:20:20 is better than the ratio of 70:15:15 and 80:10:10 in the accuracy of the test data. In this study, using 50 epochs to train the CNN model, it still resulted in an accuracy test of 80% and above. For test results the greatest accuracy is to use a ratio of 60:20:20 with an accuracy test result of 90%. The author suggests that further research can show a comparison of other methods such as ANN or RNN using Indonesian language datasets and use denoise at the preprocessing stage.

STATE ISLAMIC UNIVERSITY  
SUNAN KALIJAGA  
YOGYAKARTA

# BAB I PENDAHULUAN

## 1.1. Latar Belakang

Pengenalan suara (voice recognition) yang memungkinkan komputer mampu memproses suara atau sinyal audio menjadi teks atau melakukan hal lain. Fitur pengenalan suara dapat ditemukan pada banyak aplikasi seperti, Google OK, kecerdasan buatan milik Apple bernama Siri, webcaptioner yang merupakan aplikasi yang memproses suara menjadi teks secara langsung, dan masih banyak lagi.

Pusat Layanan Difabel (PLD) merupakan Lembaga di kampus UIN Sunan Kalijaga, guna untuk melayani mahasiswa dengan disabilitas, yakni mahasiswa tunanetra, tuna daksa, bahkan Tuli atau Tunarungu. Perkuliahan atau kegiatan yang melibatkan percakapan tentu ada kekurangan atau kurangnya akses bagi mahasiswa Tuli, yaitu mereka tidak bisa mendengar sehingga membutuhkan relawan PLD yang bersedia melakukan sebagai penerjemah atau juru Bahasa isyarat atau notulen. Setiap melakukan juru Bahasa isyarat atau notulensi tentu ada beberapa keberatan yaitu mengikuti kata-kata yang diucapkan oleh dosen atau beberapa kata asing yang belum ada isyaratnya sehingga membuat penerjemah terpaksa mengeja huruf-huruf yang memakan waktu. Oleh karena itu diharapkan dapat menggunakan fitur *voice to text* yang mungkin dapat memudahkan mahasiswa Tuli dalam perkuliahan maupun percakapan.

Dalam penelitian ini akan menggunakan metode *Convolutional Neural Network* untuk melakukan tugas klasifikasi atau pelatihan model, bagaimanapun CNN hanya bisa mengambil *input* gambar. Maka diperlukan cara mengonversi data suara menjadi *spectrogram* agar dapat digunakan sebagai *input* gambar yang akan diambil oleh CNN kemudian. *Spectrogram* menunjukkan bagaimana konten frekuensi dari sinyal berubah dari waktu ke waktu dan dapat dihitung dari waktu domain sinyal. (Cazzani. 2017) Metode alternatif lain yang tidak perlu mengonversi suara menjadi gambar adalah menggunakan metode *Mel-Spectrogram* atau



MFCC (*Mel- Frequency Cepstral Coefficients*). Namun kedua metode ini tetap membutuhkan proses konversi suara menjadi gambar. Maka dalam penelitian ini akan menggunakan cara konversi suara menjadi *spectrogram* melalui Bahasa pemrograman Python pada Google Colab.

Namun ada metode-metode yang dapat digunakan CNN adalah ANN (*Artificial Neural Network*) dan RNN (*Recurrent Neural Network*). ANN dan RNN disebut kurang kuat dari CNN. CNN dianggap lebih kuat dan lebih cepat dari ANN dan RNN. Maka CNN akan digunakan untuk penelitian ini. (GeeksforGeeks. 2020)

Agar dapat dilakukan penelitian pengenalan suara ini tentu memerlukan dataset suara, penulis mengumpulkan dataset suara berasal dari relawan-relawan PLD dengan sekitar usia di antara 19 sampai dengan 23 tahun, baik laiki-laki dan perempuan. Kata-kata yang diambil adalah inklusif, pecah, coba, dan miring sebanyak 50 data masing-masing. Alasan memilih 4 kata ini karena setiap kata ini memiliki huruf vocal dan huruf konsonan. Rata-rata durasi data suara yang diisikan oleh relawan sepanjang 1 sampai 2 detik dalam situasi lingkungan yang tenang.

Penelitian sebelumnya telah menggunakan pembelajaran mesin dengan menggunakan metode CNN untuk melakukan pengenalan suara namun banyak yang menggunakan Bahasa Inggris, sehingga penelitian ini menggunakan dataset suara dalam Bahasa Indonesia dengan metode yang sama dan dapat dilakukan Analisa seberapa akurasi.

## 1.2. Rumusan Masalah

1. Apakah sistem reknogisi suara menggunakan CNN mampu dilakukan dalam Bahasa Indonesia dengan baik?
2. Berapakah akurasi dari pengenalan suara menggunakan CNN pada dataset berbahasa Indonesia dengan 4 kata yang mewakili huruf vokal dan huruf konsonan?

### 1.3. Tujuan Penelitian

1. Mengimplementasi *Convolutional Neural Network* dalam melakukan pengenalan suara dengan dataset berbahasa Indonesia
2. Menentukan hasil akurasi dari pengenalan suara menggunakan dataset berbahasa Indonesia

### 1.4. Batasan Masalah

1. Aspek yang diteliti adalah pengenalan suara pada obyek suara jernih berbahasa Indonesia dengan menggunakan *convolutional neural network*.
2. Dataset yang digunakan adalah rekaman suara terbagi dalam 4 jenis kata, inklusif, pecah, coba, dan miring.
3. Sumber dataset yang dikumpulkan merupakan dari mahasiswa yang berstatus relawan PLD (Pusat Layanan Difabel) di antara usia 19 sampai dengan 23 tahun, baik laki-laki dan perempuan.
4. Sistem dibangun tidak menggunakan bahasa pemrograman selain bahasa *Python* dan tidak menggunakan selain Google Colaboratory.

### 1.5. Manfaat Penelitian

Penelitian ini akan memberi manfaat bagi peneliti dan pembaca untuk:

1. Mengetahui bahwa CNN mampu digunakan dengan baik untuk rekognisi suara berbahasa Indonesia.
2. Mengetahui seberapa akurasi dari dataset yang digunakan dalam proses pengenalan suara.

### 1.6. Kebaruan Penelitian

Penelitian menggunakan *Convolutional Neural Network* pernah dilakukan seperti *speech to text*, *voice commands* dengan dataset berbahasa Inggris telah dilakukan banyak sekali bahkan ada beberapa penelitian diantaranya menggunakan dataset publik. Namun penelitian yang



dilakukan oleh Mustikirini, menggunakan Bahasa Indonesia dengan dataset suara yang memiliki 4 kelas kata yang berbeda, yaitu atas, bawah, kanan, dan kiri. Namun penelitian ini menggunakan metode MFCC dan SVM untuk melakukan klasifikasi suara. (Mustikarini, 2019)



## **BAB VI PENUTUP**

### **6.1 Kesimpulan**

Dapat dilihat bahwa dari hasil penelitian ini menunjukkan rasio 60:20:20 lebih baik dari rasio 70:15:15 dan 80:10:10 dalam akurasi dari data uji. Menggunakan dataset dari 4 kata yang berbeda, inklusif, pecah, coba, dan miring dengan masing-masing memiliki 50 kata. Dalam penelitian ini menggunakan 50 *epochs* untuk melakukan pelatihan model CNN tetap membuahkan hasil tes akurasi sebesar 80% keatas. Untuk hasil tes akurasi yang paling besar adalah menggunakan rasio 60:20:20 dengan hasil tes akurasi sebesar 90%. Walaupun dalam proses *image resizing* dan konversi *spectrogram* dari audio dapat mengurangi kualitas data audio yang diproseskan untuk melatih model. Dengan adanya kekurangan pada *preprocessing* data, hasil ini mampu menunjukkan bahwa CNN dapat dilakukan dengan baik dalam rekognisi suara berbahasa Indonesia.

### **6.2 Saran**

Penelitian ini tentu ada beberapa keterbatasan, oleh karena itu penguji berharap dapat memberikan saran yang mungkin dapat dijadikan sebagai perbaikan untuk penelitian selanjutnya. Berikut ini merupakan saran-saran:

1. Penelitian sebaiknya menggunakan dataset yang lebih besar dan lebih banyak kata. Setidaknya lebih dari 100 data per kata, semakin bagus pelatihan model ini dan mungkin dapat membuahkan hasil yang lebih baik.
2. Penelitian selanjutnya dapat memberikan perbandingan metode CNN dengan metode lain seperti ANN dan/atau RNN tetapi menggunakan dataset berbahasa Indonesia.
3. Melakukan pemeriksaan konversi kembali dari

*spectrogram* ke audio guna untuk membandingkan kualitas audio orisinalnya dan audio yang telah dikonversikan.

4. Memastikan preprocessing data dengan kualitas *noise* sangat rendah agar dapat diketahui data yang mana terdeteksi berisik atau tidak.



## DAFTAR PUSTAKA

- Minkyu Lim, Donghyun Lee, Hosung Park, Yoseb Kang, Junseok Oh, Jeong-Sim Park, Gil-Jin Jang, Ji-Hwan Kim. (2018) *Convolutional Neural Network based Audio Event Classification*
- Zhang, Y., Pezeshki, M., Brakel, P., Zhang, S., Bengio, C. L. Y., & Courville, A. (2017). *Towards End-to-End Speech Recognition with Deep Convolutional Neural Networks*. <http://arxiv.org/abs/1701.02720>
- Azis, N., Herwanto, Ramadhani, F. (2021) *Implementasi Speech Recognition Pada Aplikasi E-Prescribing Menggunakan Algoritme Convolutional Neural Network*
- Mustikarini, W., Hidayat, R., & Bejo, A. (2019). Real-Time Indonesian Language Speech Recognition with MFCC Algorithms and Python-Based SVM. *IJITEE (International Journal of Information Technology and Electrical Engineering)*, 3(2), 55. <https://doi.org/10.22146/ijitee.49426>
- Palaz, A., Magimai.Doss, M., Collobert, R. (2015) *Analysis of CNN-based Speech Recognition System using Raw Speech as Input*
- Sharan, R. V., IEEE, Berkovsky, S, Liu, S. (2020) *Voice Command Recognition Using Biologically Inspired Time Frequency Representation and Convolutional Neural Networks*
- Liang, H., Lin, X., Zhang, Q., Kang, X. (2017) *Recognition of spoofed voice using Convolutional Neural Networks* <https://doi.org/10.1109/GlobalSIP.2017.8308651>
- Franti, E., Ispas, I., Dragomir, V., Dascalu, M., Zoltan. E., Stoica, I. C. (2017) *Voice Based emotion Recognition with Convolutional Neural Networks for Companion Robots*
- Pedro Dominguez-Morales, J., Wu, J., Yılmaz, E., Zhang, M., Li, H., & Tan, K. C. (2020). Deep Spiking Neural Networks for Large Vocabulary Automatic Speech Recognition. *Frontiers in Neuroscience | Www.Frontiersin.Org*, 1. <https://doi.org/10.3389/fnins.2020.00199>
- Zeyer, A., Irie, K., Schlüter, R., & Ney, H. (2018). *Improved training of end-to-end attention models for speech recognition*. <http://www.apptek.com/>,3NNAISENSE,Switzerland,<https://nnaisense.com/>

- Ardy Wicaksono<sup>1</sup> , Sisdarmanto Adinandra<sup>2</sup> , Yudi Prayudi. (2020). *Penggabungan Metode Itakura Saito Distance dan Backpropagation Neural Network untuk Peningkatan Akurasi Suara pada Audio Forensik*
- Rawat, Waseem; Wang, Zenghui (2017). *Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review*. *Neural Computation*, (), 1–98. doi:10.1162/NECO\_a\_00990
- Kubanek, Mariusz; Bobulski, Janusz; Kulawik, Joanna (2019). *A Method of Speech Coding for Speech Recognition Using a Convolutional Neural Network*. *Symmetry*, 11(9), 1185–. doi:10.3390/sym11091185
- Kembuan, O., Caren Rorimpandey, G., & Milian Tompunu Tengker, S. (2020). *Convolutional Neural Network (CNN) for Image Classification of Indonesia Sign Language Using Tensorflow*. 2020 2nd International Conference on Cybernetics and Intelligent System (ICORIS). doi:10.1109/icoris50180.2020.9
- Machine Learning Mastery, What is a Confusion in Machine Learning (2016). Retrieved March 21, 2021 from <https://machinelearningmastery.com/confusion-matrix-machine-learning/>
- Cazzani, Dario (2017). *Audio processing in Tensorflow* Retrieved April 10, 2021 from <https://towardsdatascience.com/audio-processing-in-tensorflow-208f1a4103aa>
- GeeksforGeeks, Difference between ANN, CNN and RNN (2020). Retrieved April 10, 2021 from <https://www.geeksforgeeks.org/difference-between-ann-cnn-and-rnn/>
- Computer Hope (2020). *Voice Recognition*. Retrieved April 13, 2021 from <https://www.computerhope.com/jargon/v/voicreco.htm>
- Wikipedia (2021). *Waveform*. Retrieved April 13, 2021 from <https://en.wikipedia.org/wiki/Waveform>
- Oliveira, Dimitre (2021). *A Gentle Introduction to Audio Classification With Tensorflow*. Retrieved April 15, 2021 from <https://towardsai.net/p/l/a-gentle-introduction-to-audio-classification-with-tensorflow>
- Izotope (2021). *Understanding Spectrograms*. Retrieved April 15, 2021 from <https://www.izotope.com/en/learn/understanding-spectrograms.html>
- Venturelli, Marco (2021). *The dangers behind image resizing*. Retrieved April 15, 2021 from <https://blog.zuru.tech/machine-learning/2021/08/09/the-dangers-behind-image-resizing>
- Dumane, Govinda (2020). *Introduction to Convolutional Neural Network (CNN) using Tensorflow*. Retrieved April 15, 2021 from <https://towardsdatascience.com/introduction-to-convolutional-neural-network-cnn->

de73f69c5b83



STATE ISLAMIC UNIVERSITY  
**SUNAN KALIJAGA**  
YOGYAKARTA