

**SKRIPSI**  
**SISTEM IDENTIFIKASI SUARA BERBASIS *CONVOLUTIONAL NEURAL***  
***NETWORK***



**Disusun untuk Memenuhi Syarat Mendapatkan Gelar Sarjana Teknik pada**

**Jurusan Teknik Elektro Fakultas Teknik**

**Universitas Sriwijaya**

**Oleh :**

**ALVIO YUNITA PUTRI**

**03041181520031**

**JURUSAN TEKNIK ELEKTRO**

**FAKULTAS TEKNIK**

**UNIVERSITAS SRIWIJAYA**

**2019**

**LEMBAR PENGESAHAN**  
**SISTEM IDENTIFIKASI SUARA BERBASIS *CONVOLUTIONAL NEURAL***  
***NETWORK***



**SKRIPSI**

**Disusun untuk Memenuhi Syarat Mendapatkan Gelar Sarjana Teknik pada**

**Jurusan Teknik Elektro Fakultas Teknik**

**Universitas Sriwijaya**

**Oleh :**

**ALVIO YUNITA PUTRI**

**03041181520031**

**Mengetahui,  
Ketua Jurusan Teknik Elektro**

  
**Muhammad Abu Bakar Sidik, S.T., M.Eng., Ph.D.**  
**NIP : 197108141999031005**

**Palembang, Desember 2019**  
**Menyetujui,**  
**Pembimbing Utama**

  
**Dr. Eng. Suci Dwijayanti, S.T., M.S**  
**NIP : 198407302008122001**

## HALAMAN PERNYATAAN INTEGRITAS

Yang bertanda tangan di bawah ini :

Nama : Alvio Yunita Putri  
NIM : 03041181520031  
Fakultas : Teknik  
Jurusan/Prodi : Teknik Elektro  
Universitas : Universitas Sriwijaya

Hasil Pengecekan Software *iThenticate/Turnitin* : 2 %

Menyatakan bahwa tugas akhir saya yang berjudul “Sistem Identifikasi Suara Berbasis *Convolutional Neural Network*” merupakan hasil karya sendiri dan benar keasliannya. Apabila ternyata dikemudian hari ditemukan unsur penjiplakan/plagiat dalam karya ilmiah ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya dan tanpa paksaan.

Indralaya, Desember 2019



Alvio Yunita Putri

NIM. 03041181520031

Saya sebagai Pembimbing dengan ini menyatakan bahwa saya telah membaca dan menyetujui skripsi ini dan dalam pandangan saya ruang lingkup dan kualitas skripsi ini mencukupi sebagai skripsi mahasiswa sarjana strata satu (S1).

Tanda Tangan :  \_\_\_\_\_

Pembimbing Utama : SUCI DWIJAYANTI

Tanggal : 17 / 12 / 2019

## KATA PENGANTAR

Puji dan syukur penulis ucapkan kepada Allah subhanahu wa ta'ala serta shalawat dan salam agar tercurah kepada Nabi Muhammad shallallahu 'alaihi wasallam, keluarga dan para sahabat. Berkat rahmat dan ridho Allah Subhanahu wa ta'ala, penulis dapat membuat skripsi ini yang berjudul “Sistem Identifikasi Suara Berbasis *Convolutional Neural Network*”.

Pembuatan skripsi ini sebagai salah satu syarat untuk memperoleh gelar Sarjana Teknik pada Jurusan Teknik Elektro Fakultas Teknik Universitas Sriwijaya. Dalam kesempatan ini penulis mengucapkan terima kasih kepada :

1. Bapak Muhammad Abu Bakar Sidik, S.T., M.Eng., Ph.D. selaku Ketua Jurusan Teknik Elektro dan Ibu Dr. Herlina, S.T., M.T. selaku Sekretaris Jurusan Teknik Elektro.
2. Terima kasih kepada Ibu Dr. Eng. Suci Dwijayanti, S.T., M.S. selaku pembimbing utama tugas akhir.
3. Dosen pembimbing akademik Bapak Muhammad Abu Bakar Sidik, S.T., M.Eng., Ph.D. yang telah membimbing penulis selama masa perkuliahan dan memberi saran serta masukan dalam pengambilan mata kuliah.
4. Terima kasih kepada Bapak Dr. Bhakti Yudho Suprpto, S.T., M.T.. selaku dosen yang juga ikut membimbing dalam tugas akhir ini.
5. Segenap Dosen Jurusan Teknik Elektro Fakultas Teknik Universitas Sriwijaya yang telah memberikan ilmu selama perkuliahan.
6. Orangtua, keluarga, dan saudara yang telah memberikan dukungan sepenuhnya selama pembuatan usulan proposal skripsi.
7. Teman-teman seperjuangan konsentrasi kendali dan komputer, yaitu Nur, Ega, Meydie, Azmin, Rhedo, Aldan, Abeng, Iqbal, Aldo, Qolbi, Adnan, Adit.
8. Kepada teman-teman elektro angkatan 2015 dan 2018 yang telah banyak membantu selama pembuatan skripsi ini.
9. Dan pihak-pihak yang sangat membantu dalam penulisan skripsi tugas akhir ini yang tidak dapat Penulis sebutkan satu persatu.

Penulis menyadari adanya kekurangan dalam penulisan usulan skripsi ini. Semoga skripsi ini dapat memberikan manfaat dan wawasan yang lebih luas kepada pembaca. Oleh karena itu, kritik, dan saran yang membangun sangat penulis harapkan agar dapat menjadi evaluasi yang baik dan berguna untuk perbaikan kedepannya.

Indralaya, Desember 2019



Alvio Yunita Putri

NIM. 03041181520031

**PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK  
KEPENTINGAN AKADEMIS**

Sebagai civitas akademik Universitas Sriwijaya, saya yang bertanda tangan di bawah ini:

Nama : Alvio Yunita Putri

NIM : 03041181520031

Jurusan/Prodi : Teknik Elektro

Fakultas : Teknik

Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Sriwijaya **Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty-Free Right*)** atas karya ilmiah saya yang berjudul:

**Sistem Identifikasi Suara Berbasis Convolutional Neural Network**

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Sriwijaya berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tulisan saya tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di Indralaya

Pada tanggal : Desember 2019

Yang menyatakan,



Alvio Yunita Putri

NIM. 03041181520031

## ABSTRAK

### SISTEM IDENTIFIKASI SUARA BERBASIS *CONVOLUTIONAL NEURAL NETWORK*

(Alvio Yunita Putri, 03041181520031, 2019, 33 halaman)

---

Suara manusia merupakan salah satu bentuk biometrik yang dapat digunakan untuk mengidentifikasi orang (*person identification*) dan dapat dimanfaatkan dalam bidang keamanan. Banyak metode yang telah dilakukan untuk mendapatkan identifikasi suara yang akurat. Metode tersebut harus didukung oleh faktor penting yaitu pemilihan fitur ekstraksi yang tepat karena fitur akan mempengaruhi hasil akurasi dari pengenalan suara. Pada penelitian ini digunakan *Short-Time Fourier Transform* (STFT) untuk proses ekstraksi ciri dalam bentuk spektogram. Kemudian, spektogram akan diolah oleh *Convolutional Neural Network* (CNN) sebagai metode pengidentifikasi. Data suara yang digunakan dalam penelitian ini diperoleh dari rekaman suara menggunakan *microphone*. Rekaman suara berjumlah 780 data primer yang diambil dari 78 orang mahasiswa yang terdaftar melakukan praktikum di Laboratorium Kendali dan Robotika, Jurusan Teknik Elektro Fakultas Teknik Universitas Sriwijaya. Penelitian ini menggunakan 2 tipe arsitektur CNN, yaitu arsitektur sederhana CNN dan arsitektur CNN-VGG model VGG-f. Arsitektur CNN yang dirancang pada penelitian ini menggunakan model VGG-f yang terdiri dari *convolutional layer*, *pooling layer* dan *softmaxloss* sebagai klasifikasi. Dengan parameter ukuran 224 x 224, *learning rate* 0.001 dan *batch size* 256. Hasil pengujian nilai akurasi rata-rata identifikasi suara adalah 98.7%. Hal ini menunjukkan bahwa kombinasi spektogram dan CNN dapat menghasilkan akurasi identifikasi yang baik.

***Kata kunci:*** *Convolutional Neural Network (CNN), Deep Learning, Voice Recognition, Biometric, Spektogram.*

**ABSTRACT**

**CONVOLUTIONAL NEURAL NETWORK - BASED SPEECH  
IDENTIFICATION SYSTEM**

(Alvio Yunita Putri, 03041181520031, 2019, 33 pages)

---

Voice can be utilized as person identification in a biometric system which can be implemented in a security system. Various methods have been proposed to obtain accurate speech identification. The accuracy of such methods were determined by feature extraction. Hence, this study proposes to utilize raw feature. Namely spectrogram which is voice image representation obtained from Short-Time Fourier Transform (STFT). Later, the spectrogram for each sample is processed by Convolutional Neural Network (CNN) as an identifier machine. In this research, the voice is recorded using a microphone. Those data are primary data obtained from 78 students enrolled in the Laboratory of Control and Robotics, Departement of Electrical Engineering, Faculty of Engineering, Sriwijaya University. Each individu records his/her voice 10 times. Thus, there are 780 data in total. Next, 2 types of CNN architecture are utilized, namely simple-CNN architecture and VGG-f architecture. The CNN architecture designed in this study uses the VGG-f model consist of the convolutional layer, pooling layer and softmaxloss as a classification with a parameter size of  $224 \times 224$ , a learning rate of 0.001 and a batch size of 256. The results show that the accuracy is 98.7%. It may imply that the combination of spectrogram and CNN may improve the accuracy of speech identification.

**Keywords : Convolutional Neural Network (CNN), Deep Learning, Voice Recognition, Biometric, Spectrogram**

.

# DAFTAR ISI

|   | Halaman |
|---|---------|
| <b>HALAMAN JUDUL</b> .....                            | i       |
| <b>ABSTRAK</b> .....                                  | ii      |
| <b>DAFTAR ISI</b> .....                               | x       |
| <b>DAFTAR GAMBAR</b> .....                            | xii     |
| <b>DAFTAR TABEL</b> .....                             | xiii    |
| <b>DAFTAR LAMPIRAN</b> .....                          | xiv     |
| <br><b>BAB I PENDAHULUAN</b>                          |         |
| 1.1. Latar Belakang .....                             | 1       |
| 1.2. Perumusan Masalah .....                          | 3       |
| 1.3. Tujuan Penulisan .....                           | 4       |
| 1.4. Pembatasan Masalah .....                         | 4       |
| 1.5. Keaslian Penulisan .....                         | 4       |
| <br><b>BAB II TINJAUAN PUSTAKA</b>                    |         |
| 2.1. <i>State Of The Art</i> .....                    | 7       |
| 2.2. Pengenalan Suara .....                           | 10      |
| 2.3. Proses <i>Pre-processing</i> .....               | 11      |
| 2.4. <i>Short-Time Fourier Transform (STFT)</i> ..... | 12      |
| 2.5. <i>Convolutional Neural Network (CNN)</i> .....  | 14      |
| 2.6. Perangkat Keras dan Perangkat Lunak .....        | 18      |
| 2.6.1. Perangkat Keras .....                          | 18      |
| 2.6.2. Perangkat Lunak .....                          | 19      |

### **BAB III METODOLOGI PENELITIAN**

|                               |    |
|-------------------------------|----|
| 3.1. Studi Literatur .....    | 21 |
| 3.2. Pengambilan Data .....   | 21 |
| 3.3. Perancangan Sistem ..... | 22 |
| 3.4. Pengujian .....          | 26 |

### **BAB IV HASIL DAN PEMBAHASAN**

|  |    |
|--|----|
| 4.1. Pengolahan Data dan <i>Training Network</i> ..... | 27 |
| 4.2. <i>Training</i> Data dan Pengujian.....           | 28 |
| 4.3. Model Hasil Pelatihan.....                        | 29 |
| 4.4. Pengujian Model Hasil Pelatihan .....             | 32 |
| 4.5. Analisa Hasil Pengujian .....                     | 34 |

### **BAB V KESIMPULAN DAN SARAN**

|                       |    |
|-----------------------|----|
| 5.1. Kesimpulan ..... | 35 |
| 2.2. Saran .....      | 35 |

### **DAFTAR PUSTAKA**

### **LAMPIRAN**

## DAFTAR GAMBAR

| Gambar   | Halaman |
|--|---------|
| <b>Gambar 2.1.</b> Perbandingan dari tingkat akurasi dialek .....                  | 7       |
| <b>Gambar 2.2.</b> Spectrogram (diatas) dan Full-Mel Spectrogram (dibawah) .....   | 8       |
| <b>Gambar 2.3.</b> Perbandingan bobot filter setelah pembelajaran .....            | 9       |
| <b>Gambar 2.4.</b> Blok diagram pembelajaran pola dan pengenalan suara .....       | 10      |
| <b>Gambar 2.5.</b> Contoh Jaringan CNN.....  | 14      |
| <b>Gambar 2.6.</b> Lapisan pada CNN.....   | 15      |
| <b>Gambar 2.7.</b> Proses Konvolusi .....  | 16      |
| <b>Gambar 2.8.</b> Fungsi ReLU .....   | 16      |
| <b>Gambar 2.9.</b> Operasi <i>Max Pooling</i> .....                                | 17      |
| <b>Gambar 2.10.</b> Perangkat perekam suara .....                                  | 18      |
| <b>Gambar 2.11.</b> USB <i>Sound Card</i> .....                                    | 19      |
| <b>Gambar 2.12.</b> <i>Software Audacity</i> .....                                 | 19      |
| <b>Gambar 2.13.</b> Perangkat Lunak Matlab .....                                   | 20      |
| <b>Gambar 3.1.</b> <i>Flowchat</i> Perancangan Sistem.....                         | 20      |
| <b>Gambar 3.2.</b> Tampilan data rekaman .....                                     | 21      |
| <b>Gambar 3.3.</b> Proses Ekstraksi Ciri STFT .....                                | 22      |
| <b>Gambar 3.3.</b> Tampilan <i>spectrogram</i> setelah proses ekstraksi ciri ..... | 23      |
| <b>Gambar 4.1.</b> Data latih <i>spectrogram</i> berukuran 875x656 piksel .....    | 25      |
| <b>Gambar 4.2.</b> Plot grafik pelatihan awal 100 epoch dengan CNN .....           | 27      |
| <b>Gambar 4.3.</b> Plot grafik pelatihan awal 100 epoch dengan model VGG-f.....    | 29      |

## DAFTAR TABEL

| Tabel   | Halaman |
|---|---------|
| <b>Tabel 4.1.</b> Parameter pelatihan data spektrogram .....                      | 26      |
| <b>Tabel 4.2.</b> Arsitektur CNN .....  | 26      |
| <b>Tabel 4.3.</b> Parameter <i>training network</i> model VGG-f .....             | 28      |
| <b>Tabel 4.4.</b> Arsitektur Model VGG-f.....                                     | 29      |
| <b>Tabel 4.5.</b> Sampel Akurasi Data <i>Testing</i> Pengenalan Spektrogram ..... | 30      |

## DAFTAR LAMPIRAN

- Lampiran I Hasil Pengujian Pengenalan Spektogram  
Lampiran II Lembar Persentase Plagiarisme

## **BAB I**

### **PENDAHULUAN**

#### **1.1. Latar Belakang**

Perkembangan teknologi yang pesat memberikan dampak positif dan negatif bagi kehidupan manusia. Salah satu dampak negatif yang menjadi perhatian adalah masalah sistem keamanan data. Sistem keamanan data dapat disalahgunakan untuk melakukan kejahatan seperti pencurian, perusakan, dan penyalahgunaan data yang tersimpan. Untuk memecahkan permasalahan tersebut diperlukan suatu sistem keamanan yang memungkinkan seseorang hanya dapat dikenali dengan karakteristik yang dimilikinya. Beberapa contoh sistem keamanan data berdasarkan identifikasi seseorang, antara lain: *pattern lock*, *pin lock*, *password lock*, dan sebagainya. Namun, metode tersebut memiliki kelemahan karena mudah untuk ditiru atau dicuri orang yang tidak berkepentingan. Salah satu cara untuk mengatasi hal tersebut adalah menggunakan karakteristik alami manusia sebagai input pada sistem keamanan. Pengenalan identitas seseorang menggunakan karakteristik alami manusia sebagai basisnya dikenal dengan istilah biometrik. Pengenalan biometrik akan menggantikan sistem keamanan kata sandi dan kunci dikarenakan permintaan sistem keamanan yang semakin meningkat [1]. Pengenalan melalui biometrik ini memenuhi dua fungsi penting yaitu identifikasi dan verifikasi. Sistem identifikasi bertujuan untuk menentukan identitas seseorang. Sedangkan sistem verifikasi bertujuan untuk menolak dan menerima identitas yang diklaim seseorang. Cara kerja teknologi biometrik yaitu dengan menggunakan teknik *pattern recognition* atau teknik pengenalan pola. Ada berbagai macam pola yang dapat dikenali seperti sidik jari, garis telapak tangan, wajah, iris mata dan pengenalan suara.

Suara manusia merupakan salah satu bentuk biometrik yang dapat digunakan untuk mengidentifikasi orang (*person identification*). Dibandingkan dengan pengenalan biometrik lainnya, keuntungan *Speech Recognition* (SR) diantaranya adalah tidak memerlukan biaya yang besar karena tidak menggunakan

hardware khusus dan ukuran filenya yang kecil [1]. Suatu algoritma yang mengubah ucapan manusia menjadi teks yang dimaksudkan oleh pembicara adalah *Automatic Speech Recognition* (ASR). Banyak metode yang telah dilakukan untuk mendapatkan pengenalan suara yang akurat. Penelitian yang dilakukan oleh Ayu Mawadda Warohma dkk membahas tentang identifikasi dialek yang ada di Indonesia dengan menggunakan *Mel Frequency Cepstrum Coefficient* (MFCC) dan Jaringan Syaraf Tiruan (JST) [2]. Namun, penelitian ini masih memiliki kelemahan pada *feature* MFCC yang digunakan karena MFCC dapat menyebabkan hilangnya informasi suara asli akibat proses pengurangan atau reduksi dimensi. Penelitian yang lain menggunakan metode *Hidden Markov Models* (HMM) melalui estimasi awal dan kemudian estimasi berulang-ulang dengan menggunakan gradien (*hill-climbing*), yang hanya menjamin model yang optimal secara local [3]. Kesulitan utama dengan HMM adalah dari penggunaan model Markov orde pertama. Salah satu cara untuk membantu proses tersebut adalah penggunaan "*Delta*" dan "*Delta-delta*" untuk memasukkan informasi waktu pada beberapa frames tetapi cara ini tidak efisien karena harus memanfaatkan waktu dengan lebih baik.

Pada penelitian tentang *Speech Recognition*, salah satu faktor penting yaitu pemilihan fitur ekstraksi yang baik karena fitur akan mempengaruhi hasil akurasi dari pengenalan suara. Penelitian tentang fitur ekstraksi tersebut telah dilakukan oleh beberapa peneliti. Salah satunya adalah Ruchao Fan dkk yang membahas perbandingan fitur *input*. Penelitian ini mengusulkan penggunaan *Full-Mel Spectrogram feature* (FBANK) sebagai fitur ekstraksi. FBANK ini dipilih karena struktur spektrum pada frekuensi rendah lebih jelas, dan lebih jarang pada frekuensi tinggi yang sesuai dengan karakteristik persepsi nonlinear untuk telinga manusia[4]. Namun, penelitian ini memiliki kekurangan yaitu proses komputasi yang sangat kompleks. Untuk mengatasi kelemahan FBANK, penulis mencoba menerapkan algoritma *Short-Time Fourier Transform* (STFT). STFT memiliki keuntungan dibandingkan dengan *Fast Fourier Transform* (FFT) yang hanya akan mendapatkan informasi frekuensi dengan cukup akurat namun tanpa informasi waktu dari suatu sinyal dan sebaliknya. STFT dapat menjembatani dua hal tersebut,

pada saat bersamaan STFT akan memberikan informasi waktu dan frekuensi secara spesifik. Sehingga, penulis akan menggunakan *raw feature* (spectrogram) sebagai fitur ekstraksi pada *Speech Recognition*.

Pada penelitian diatas, metode pengenalan yang digunakan masih menggunakan metode tradisional seperti HMM. Metode ini memiliki kelemahan terutama apabila suara yang digunakan terdistorsi oleh *noise*. Kualitas suara dapat terganggu apabila lingkungan sumber suara banyak mengandung *noise* yang melatarbelakangi informasi suara tersebut. Rekaman suara yang terdistorsi *noise* menyebabkan terganggunya proses identifikasi. Oleh karena itu, pada penelitian ini, penulis akan menggunakan metode *Convolutional Neural Network* (CNN). CNN sering digunakan untuk mengenali benda atau citra sehingga untuk pengenalan suara metode CNN belum banyak dilakukan. Tomonori Kawamura dkk melakukan penelitian yaitu mengidentifikasi suara iringan dan suara nyanyian menggunakan metode CNN [5]. Pada penelitian ini, CNN menunjukkan hasil yang lebih baik ketika proses dari *raw waveform* dan *filter learning* digabungkan. Sehingga, penelitian ini dapat dikatakan memiliki fitur yang efektif untuk memperkirakan FO (*Fundamental Frequency*).

Dari semua pembahasan di atas, penulis akan menggunakan metode CNN sebagai *recognizer* untuk mengidentifikasi identitas seseorang dengan fitur ekstraksi menggunakan metode STFT sehingga menghasilkan tingkat akurasi yang baik untuk sistem keamanan berbasis biometrik suara.

## 1.2. Perumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, penulis mengambil rumusan masalah dari penelitian-penelitian sebelumnya yaitu penggunaan algoritma. Algoritma pada penelitian sebelumnya membuat proses komputasi yang sangat kompleks sehingga untuk mengidentifikasi suara yang telah terdistorsi *noise* membutuhkan waktu yang lama. Selain itu, pemilihan fitur ekstraksi merupakan faktor yang sangat penting untuk mempengaruhi hasil akurasi. Pada penelitian ini, fitur ekstraksi yang digunakan adalah *raw feature* (spectrogram). Kemudian,

metode yang akan digunakan untuk identifikasi suara pada sistem keamanan ini adalah CNN karena masih belum banyak diimplementasikan.

### **1.3. Tujuan Penulisan**

Penulisan tugas akhir ini bertujuan untuk :

Mengembangkan suatu sistem berbasis CNN yang akan digunakan dalam mengidentifikasi suara seseorang dan Menunjukkan performansi dari sistem identifikasi suara dengan metode CNN.

### **1.4. Pembatasan Masalah**

Beberapa batasan perlu diberikan agar permasalahan yang dibahas pada tugas akhir ini menjadi terarah, yaitu :

1. Perekaman menggunakan Rode Video Micro Microphone dengan aplikasi Audacity.
2. Ekstraksi ciri yang digunakan *Short-Time Fourier Transform* (STFT).
3. Metode pelatihan dan pengambilan keputusan menggunakan algoritma *Convolutional Neural Network* (CNN).
4. Data uji terdiri dari 78 sampel suara. Dengan 10 suara 3 digit terakhir setiap NIM Mahasiswa.
5. Frekuensi *sampling* yang digunakan 16 kHz.
6. Proses perekaman dilakukan di Laboratorium Kendali dan Robotika.
7. File perekam suara disimpan pada format wav.
8. Perancangan program menggunakan Software Matlab R2018a.

### **1.5. Keaslian Penelitian**

Penelitian yang dilakukan ini berdasarkan penelitian-penelitian sebelumnya, yaitu Ayu Mawadda Warohma dkk membahas tentang identifikasi dialek yang ada di Indonesia dengan menggunakan MFCC (*Mel Frequency Cepstrum Coefficient*) dan *Neural Network* [2]. Penelitian yang dilakukan terdiri dari 3 dialek di Indonesia, yaitu dialek Jawa, Minang dan Batak. Tingkat akurasi

yang lebih tinggi pada dialek orang Jawa dan Minang daripada orang Batak. Sehingga penelitian ini menunjukkan bahwa karakteristik dialek dapat mempengaruhi *recognition*. Tingkat keberhasilan juga dipengaruhi oleh jumlah neuron. Semakin rumit model pengenalan, semakin banyak jumlah neuron yang dibutuhkan. Dengan menggunakan metode MFCC, metode ini telah berhasil mengidentifikasi dialek daerah dan identifikasi yang paling tepat adalah dialek Minang dan Jawa. Namun, penelitian ini masih memiliki kelemahan pada *Feature* MFCC yang digunakan. MFCC dapat menyebabkan hilangnya informasi suara asli akibat proses pengurangan dimensi.

Fan dan Gang Liu membahas tentang penggunaan CNN pada pemilihan fitur input dan desain parameter CNN [4]. Dalam perbandingan fitur input, makalah ini mengusulkan FBANK (*Full-Mel Spectrogram feature*) karena didapatkan bahwa struktur spektrum pada frekuensi rendah lebih jelas dan lebih jarang pada frekuensi tinggi yang sejalan dengan karakteristik persepsi nonlinear untuk telinga manusia. Menurut jurnal ini, kelemahan dalam menggunakan CNN apabila kekurangan data maka sejumlah besar parameter di CNN akan menyebabkan pemasangan yang berlebihan. Selain itu, susahnya untuk mengganti model bahasa misalnya, bahasa n-gram dengan model bahasa RNN serta bagaimana cara melatih bersama-sama model bahasa dan model akustik.

Tomonori Kawamura, Atsuhiko Kai dan Seiichi Nakagawa membahas tentang identifikasi suara iringan dan suara nyanyian dengan menggunakan beberapa metode, yaitu 1. RAPT dan DNN, 2. DNN dan 3. CNN [5]. Selain itu, adanya perbandingan layer dengan menggunakan *filter learning* dan tidak menggunakan *filter learning*. Kelebihan dari jurnal penelitian ini adalah penjelasannya sangat lengkap, berdasarkan hasil penelitiannya tingkat kebenarannya bernilai bagus saat menggunakan *filter learning* dan menggunakan Metode kombinasi antara CNN dan DNN dan dengan menggunakan pembelajaran CNN suara *noise* dapat di estimasi. Sedangkan, Kelemahan dari jurnal penelitian ini adalah suara iringan dan suara nyanyian hanya berbahasa China dan kurangnya peningkatkan akurasi pada pengenalan suara.

Dari semua penelitian diatas, penulis akan menggunakan metode CNN sebagai *recognizer* untuk mengidentifikasi identitas seseorang dengan Fitur Ekstraksi menggunakan metode STFT sehingga menghasilkan tingkat akurasi yang baik.

## DAFTAR PUSTAKA

- [1] I. A. Mahmudi, Suwandi, and H. Bethaningtyas, “3 1,2,3,” vol. 2, no. 2, pp. 3223–3232, 2015.
- [2] A. M. Warohma, “Identification of Regional Dialects Using Mel Frequency Cepstral Coefficients ( MFCCs ) and Neural Network,” *2018 Int. Semin. Appl. Technol. Inf. Commun.*, pp. 522–527, 2018.
- [3] D. O. Shaughnessy, “Automatic speech recognition,” pp. 417–424, 2015.
- [4] R. Fan and G. Liu, “CNN-Based Audio Front End Processing on Speech Recognition,” *2018 Int. Conf. Audio, Lang. Image Process.*, pp. 349–354, 2018.
- [5] T. Kawamura and A. Kai, “Noise Robust Fundamental Frequency Estimation of Speech using CNN-based discriminative modeling,” *IEEE*, 2018.
- [6] H. N. M. Shah, M. Z. Ab. Rashid, M.F. Abdollah, M. N. Kamarudin, C.K. Lin, and Z. Kamis, “Biometric Voice Recognition in Security System,” *Indian Journal of Science and Technology*, vol 7, no. 2, Februari., pp. 104–112, 2014. [E-PDF].
- [7] A. Setiawan, A. Hidayatno, R. R. Isnanto, J. T. Elektro, F. Teknik, and U. Diponegoro, “Angga Setiawan\*, Achmad Hidayatno\*\*, R. Rizal Isnanto\*\* Jurusan Teknik Elektro, Fakultas Teknik, Universitas Diponegoro, Jln. Prof. Sudharto, Tembalang, Semarang, Indonesia,” pp. 1–7.
- [8] J. Liu and X. Xiang, “Review of the anti-noise method in the speech recognition technology,” pp. 1391–1394, 2017.
- [9] M. Parchami, W. Zhu, B. Champagne, and E. Plourde, “Recent Developments in Speech Enhancement in the Short-Time Fourier Transform Domain,” pp. 45–77, 2016.
- [10] D. Anggraeni, W. S. M. Sanjaya, M. Munawwaroh, M. Y. S. Nurasyidiek, and I. P. Santika, “Control of Robot Arm based on Speech Recognition using Mel-Frequency Cepstrum Coefficients ( MFCC ) Method,” *2017 Int. Conf. Adv. Mechatronics, Intell. Manuf. Ind. Autom.*, no. 1, pp. 217–222, 2017.

- [11] O. Abdel-hamid, A. Mohamed, H. Jiang, L. Deng, G. Penn, and D. Yu, "Convolutional Neural Networks for Speech Recognition," vol. 22, no. 10, pp. 1533–1545, 2014.
- [12] S. Samuel, "Pengenalan Deep Learning Part 7 : Convolutional Neural Network (CNN)," 2017. [Online]. Available: <https://medium.com/@samuelsena/pengenalan-deep-learning-part-7-convolutional-neural-network-cnn-b003b477dc94>. [Accessed: 25-Feb-2019].
- [13] M. Wang, Z. Wang, and J. Li, "Deep Convolutional Neural Network Applies to Face Recognition in Small and Medium Databases," in *11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, China, no. Icsai, pp. 1368–1372, 2018.
- [14] P. Wozniak, H. Afrisal, R. G. Esparza, and B. Kwolek, "Scene recognition for indoor localization of mobile robots using deep CNN," in *International Conference on Computer Vision and Graphics (ICCVG)*, Poland, September, 2018.
- [15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Network," *Neural Information Processing System*, vol.25, Jan, 2012.