

**KLASIFIKASI GOLONGAN SUARA DALAM PADUAN SUARA  
DENGAN MENGGUNAKAN  
CONVOLUTIONAL RECURRENT NEURAL NETWORK (CRNN)**

**Laporan Tugas Akhir**

**Disusun sebagai syarat kelulusan tingkat sarjana**

**Oleh**

**STEFANUS**

**NIM : 13519101**



**PROGRAM STUDI TEKNIK INFORMATIKA  
SEKOLAH TEKNIK ELEKTRO DAN INFORMATIKA  
INSTITUT TEKNOLOGI BANDUNG**

**Juni 2023**

**KLASIFIKASI GOLONGAN SUARA DALAM PADUAN SUARA  
DENGAN MENGGUNAKAN  
CONVOLUTIONAL RECURRENT NEURAL NETWORK (CRNN)**

**Laporan Tugas Akhir**

**Oleh**

**STEFANUS**

**NIM : 13519101**

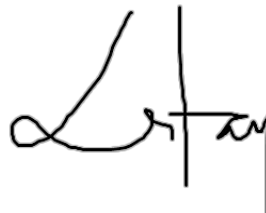
**Program Studi Teknik Informatika**

Sekolah Teknik Elektro dan Informatika

Institut Teknologi Bandung

Telah disetujui dan disahkan sebagai Laporan Tugas Akhir  
di Bandung, pada tanggal 5 Mei 2023

Pembimbing,



Dessi Puji Lestari, S.T, M.Eng., Ph.D.

NIP 197912012012122005

## LEMBAR PERNYATAAN

Dengan ini saya menyatakan bahwa:

1. Pengerjaan dan penulisan Laporan Tugas Akhir ini dilakukan tanpa menggunakan bantuan yang tidak dibenarkan.
2. Segala bentuk kutipan dan acuan terhadap tulisan orang lain yang digunakan di dalam penyusunan laporan tugas akhir ini telah dituliskan dengan baik dan benar.
3. Laporan Tugas Akhir ini belum pernah diajukan pada program pendidikan di perguruan tinggi mana pun.

Jika terbukti melanggar hal-hal di atas, saya bersedia dikenakan sanksi sesuai dengan Peraturan Akademik dan Kemahasiswaan Institut Teknologi Bandung bagian Penegakan Norma Akademik dan Kemahasiswaan khususnya Pasal 2.1 dan Pasal 2.2.

Bandung, 6 Juni 2023



Stefanus

NIM 13519101

## **ABSTRAK**

# **KLASIFIKASI GOLONGAN SUARA DALAM PADUAN SUARA DENGAN MENGGUNAKAN CONVOLUTIONAL RECURRENT NEURAL NETWORK (CRNN)**

Oleh

STEFANUS

NIM : 13519101

Penentuan golongan suara merupakan hal penting yang menjadi pondasi dari paduan suara. Penentuan golongan suara ini memerlukan pengetahuan mendalam di bidang musik. Hal ini menjadi penyebab pembangunan model pengklasifikasian golongan suara dalam paduan suara merupakan eksperimen multidisiplin. Selain peninjauan dari aspek teknis, diperlukan juga peninjauan dari aspek musikalitas yang mendalam.

Penelitian ini berfokus kepada pengklasifikasian golongan suara menggunakan fitur-fitur selain nada dasar ( $f_0$ ) seperti warna suara dan kestabilan suara. Dalam penelitian ini, data diambil dari anggota Paduan Suara Mahasiswa ITB (PSM-ITB) dengan pelabelan yang dilakukan oleh tim pelatihan Paduan Suara Mahasiswa ITB (PSM-ITB).

Penelitian ini memberikan hasil yang relatif baik, yakni akurasi 0.87 untuk model *Convolutional Recurrent Neural Network*, lebih tinggi daripada model *Convolutional Neural Network* dengan akurasi 0.74 dan *Recurrent Neural Network* dengan akurasi 0.65.

Kata kunci: paduan suara, *timbre*, MFCC.

## KATA PENGANTAR

Puji dan syukur penulis hantarkan kepada Tuhan Yang Maha Esa sebab oleh rahmat dan bimbinganNya-lah, penulis mampu menyelesaikan Tugas Akhir ini dengan baik. Tentu proses Tugas Akhir ini tidak mungkin berjalan dengan lancar tanpa ada bantuan dari berbagai pihak yang diberikan kepada penulis. Oleh sebab itu, kata pengantar ini penulis dedikasikan kepada pihak-pihak yang telah membantu penulis dalam mengerjakan Tugas Akhir ini, khususnya kepada:

1. Ibu Dessi Puji Lestari, S.T, M.Eng., Ph.D. selaku dosen pembimbing Tugas Akhir penulis yang telah membimbing penulis dari awal pembuatan Tugas Akhir ini hingga tahap akhir penulisan baik dari segi penulisan, teknis, penentuan arah, hingga motivasi dan nasihat yang diberikan selama ini.
2. Ibu Dr. Fariska Zakhralativa Ruskanda, S.T., M.T. selaku dosen wali penulis yang telah membimbing penulis sejak awal perjalanan penulis dalam lika-liku jurusan Teknik Informatika hingga akhir perjalanan studi S1 ini.
3. Dosen mata kuliah Tugas Akhir I dan II karena sudah membuat keseluruhan proses Tugas Akhir I dan II berjalan dengan baik, mulai dari pemaparan Tugas Akhir, hingga mengatur koordinasi antara dosen pembimbing, dosen penguji, dan mahasiswa.
4. Seluruh dosen pengajar yang telah mengajar penulis secara langsung maupun tidak langsung. Tempaan pengetahuan serta pengalaman para dosen pengajar-lah yang membentuk penulis sebagai pribadi seorang informatikawan.
5. Bapak Jusuf Lamlo dan Ibu Siu Tjhai selaku orang tua dari penulis, Stella Lamlo selaku kakak dari penulis, serta seluruh keluarga dari penulis yang senantiasa memberikan dukungan moral serta sumber semangat bagi penulis selama pengerjaan Tugas akhir ini.

6. Teman-teman terdekat, Jeanne D'Arc Amara Hanieka dan Giovani Anggasta, atas persahabatan, dukungan, serta momen berharga selama 4 tahun terakhir.
7. Teman-teman anggota PSM-ITB dari berbagai angkatan termasuk juga alumni, yang selain telah menjadi sumber data dalam pengerjaan Tugas Akhir ini, juga menjadi rumah, tempat berbagi canda tawa, serta tempat penulis mengaktualisasikan diri.
8. Seluruh pihak yang telah membantu baik secara langsung maupun secara tidak langsung dan tidak dapat penulis nyatakan satu persatu.

Kesempurnaan hanyalah angan, yang dapat dilakukan oleh manusia hanyalah berusaha melakukan yang terbaik, dan terus berusaha lebih baik lagi. Oleh sebab itu penulis sangat terbuka terhadap segala kritik, saran, dan masukan sebagai pembelajaran untuk kedepannya. Semoga laporan ini dapat bermanfaat bagi segala pihak yang membacanya.

## DAFTAR ISI

<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
I.1    Latar Belakang.....	1
I.2    Rumusan Masalah.....	4
I.3    Tujuan.....	4
I.4    Batasan Masalah .....	4
I.5    Metodologi.....	5
I.6    Sistematika Pembahasan.....	6
<b>BAB II STUDI LITERATUR .....</b>	<b>8</b>
II.1    Karakteristik Suara Manusia.....	8
II.1.1    Golongan Suara dalam Paduan Suara .....	8
II.1.2 <i>Vocal overexertion</i> .....	9
II.2    Data Audio.....	9
II.2.1    Sumber Data Audio.....	10
II.2.2    Fitur Audio .....	11
II.3    Metode Klasifikasi Golongan Suara.....	13
II.3.1    Ekstraksi Fitur .....	13
II.3.2    Model Klasifikasi .....	14
II.3.3    Metrik Evaluasi .....	17
II.4    Hasil Penelitian Terkait .....	18
II.5    Kakas Pengembangan.....	19
II.5.1    Kakas Praproses Data.....	20

II.5.2	Kakas Ekstraksi Fitur .....	20
II.5.3	Kakas Pemodelan .....	20
<b>BAB III</b>	<b>PENGEMBANGAN MODEL PEMBELAJARAN MESIN</b>	
	<b>PENGKLASIFIKASIAN GOLONGAN SUARA .....</b>	<b>21</b>
III.1	Analisis Persoalan .....	21
III.2	Analisis Solusi .....	22
III.3	Rancangan Solusi .....	23
III.4	Pengumpulan <i>Dataset</i> .....	25
III.5	Pengolahan Data.....	25
III.6	Pemodelan Pengklasifikasian Golongan Suara .....	26
III.7	Desain Eksperimen Model Pengklasifikasian Golongan Suara .....	27
III.8	Pengujian .....	27
<b>BAB IV</b>	<b>EVALUASI MODEL PEMBELAJARAN MESIN</b>	
	<b>PENGKLASIFIKASIAN GOLONGAN SUARA .....</b>	<b>28</b>
IV.1	Evaluasi Tahap Training .....	28
IV.2	Evaluasi Objektif Tahap Testing .....	39
IV.3	Evaluasi Subjektif Tahap Testing.....	40
IV.4	Analisis Hasil Evaluasi Tahap Testing.....	45
<b>BAB V</b>	<b>KESIMPULAN DAN SARAN .....</b>	<b>46</b>
V.1	Kesimpulan .....	46
V.2	Saran .....	46



## **DAFTAR LAMPIRAN**

<b>Lampiran A. Dataset.....</b>	<b>50</b>
<b>Lampiran B. Hasil Prediksi Data Uji .....</b>	<b>51</b>

## DAFTAR GAMBAR

Gambar I-1 6 (Enam) Bar pertama dari Partitur Mars ITB.....	2
Gambar II-1 <i>Mel-Frequency Cepstral Coefficient</i> .....	12
Gambar II-2 <i>Convolutional Neural Network</i> .....	14
Gambar II-3 <i>Recurrent Neural Network</i> .....	15
Gambar III-1 Arsitektur <i>baseline</i> CNN (kiri) dan RNN (kanan).....	23
Gambar III-2 Arsitektur <i>Convolutional Recurrent Neural Network</i> .....	24
Gambar IV-1 Titik Seimbang model CRNN dengan LR 0.001 .....	29
Gambar IV-2 Titik Seimbang Model CRNN dengan LR 0.1 .....	29
Gambar IV-3 Titik Seimbang Model CRNN dengan LR 0.0005 .....	30
Gambar IV-4 Titik Seimbang Model CRNN dengan LR 0.00001 .....	30
Gambar IV-5 Titik Seimbang Model CRNN dengan LR 0.00005 .....	31
Gambar IV-6 Titik Seimbang Model RNN dengan LR 0.0001 .....	31
Gambar IV-7 Titik Seimbang Model RNN dengan LR 0.001 .....	32
Gambar IV-8 Titik Seimbang Model RNN dengan LR 0.1 .....	32
Gambar IV-9 Titik Seimbang Model RNN dengan LR 0.0005 .....	33
Gambar IV-10 Titik Seimbang Model RNN dengan LR 0.00001 .....	33
Gambar IV-11 Titik Seimbang Model RNN dengan LR 0.00005 .....	34
Gambar IV-12 Titik Seimbang Model CNN dengan LR 0.0001 .....	34
Gambar IV-13 Titik Seimbang Model CNN dengan LR 0.001 .....	35
Gambar IV-14 Titik Seimbang Model CNN dengan LR 0.1 .....	35
Gambar IV-15 Titik Seimbang Model CNN dengan LR 0.0005 .....	36

Gambar IV-16 Titik Seimbang Model CNN dengan LR 0.00001 .....	36
Gambar IV-17 Titik Seimbang Model CNN dengan LR 0.00005 .....	37
Gambar IV-18 Titik Seimbang Model CRNN dengan LR 0.0001 .....	37
Gambar IV-19 MFCC alto-02.....	41
Gambar IV-20 MFCC alto-18.....	41
Gambar IV-21 MFCC alto-21 .....	42
Gambar IV-22 MFCC bass-21 .....	42
Gambar IV-23 MFCC bass-28.....	43
Gambar IV-24 MFCC sopran-18 .....	43
Gambar IV-25 MFCC sopran-24 .....	44
Gambar IV-26 MFCC tenor-15.....	44

## DAFTAR TABEL

Tabel II-1. <i>Vocal Range</i> untuk Tiap Golongan Suara berdasarkan The Concise Oxford Dictionary of Music (2007) .....	9
Tabel II-2 <i>Confusion Matrix</i> .....	17
Tabel IV-1 <i>Epoch</i> dengan <i>loss</i> optimal.....	38
Tabel IV-2 <i>Accuracy</i> dan <i>val_accuracy</i> untuk masing-masing <i>epoch</i> optimal ....	38
Tabel IV-3 <i>Classification Report CRNN</i> .....	39
Tabel IV-4 <i>Classification Report CNN</i> .....	40
Tabel IV-5 <i>Classification Report RNN</i> .....	40

# **BAB I**

## **PENDAHULUAN**

Bab Pendahuluan ini berisikan penjelasan atas landasan pembuatan Tugas Akhir mengenai klasifikasi golongan suara dalam paduan suara dengan menggunakan *Convolutional Recurrent Neural Network*. Bab ini terdiri dari latar belakang pelaksanaan tugas akhir, rumusan masalah, tujuan tugas akhir, batasan masalah, metodologi, serta sistematika pembahasan laporan tugas akhir.

### **I.1 Latar Belakang**

Musik adalah bahasa universal yang dapat dimengerti kendati terdapat batasan bahasa, budaya, maupun selera. Musik mampu mengantarkan pesan kepada pendengar dan penikmat musik itu sendiri. Paduan suara merupakan salah satu jenis musik paling tua yang ditemukan bukti keberadaannya sejak zaman Yunani Kuno. (Ratmono, 1985). Di Indonesia pun, paduan suara sudah cukup berkembang pesat. Bukan satu-dua lagi kompetisi dan penghargaan yang didapat oleh paduan suara di Indonesia. Paduan Suara Mahasiswa ITB (PSM-ITB) merupakan Paduan Suara Mahasiswa tertua yang ada di Indonesia.

Paduan suara mempunyai pesona tersendiri yang tidak dapat didapatkan dari jenis musik lainnya. Perbedaan paduan suara dengan jenis musik lainnya ada pada kata ‘Padu’, yang memiliki konsep berbeda dari sekadar bernyanyi bersama-sama. Salah satu aspek dari kata ‘Padu’ ini adalah Polifoni. Polifoni adalah konsep musik dimana di satu waktu, sumber musik tidak hanya satu, melainkan beberapa sumber musik yang secara paralel membunyikan nadanya masing-masing dan menciptakan harmoni yang indah. Dalam paduan suara, polifoni ini umumnya direalisasikan dengan 4 golongan suara utama, yakni Sopran, Alto, Tenor, dan Bass yang biasa disingkat dengan SATB (Ratmono, 1985).

## MARS ITB

Lagu & Syair : Drs. Ahmad Stiawan  
Arr. Sudjoko

Ala Marcia (ca.  $\text{♩} = 134$ )  
Gagah, bersemangat, staccato

Soprano  
Alto  
Tenor  
Bass  
Piano

De-rap - kan lang-kah, ta-tap ke de-pan! I - T-  
De-rap - kan lang-kah, ta-tap ke de-pan! I - T-  
De-rap - kan lang-kah, ta-tap ke de-pan! I - T-  
De-rap - kan lang-kah, ta-tap ke de-pan! I - T-

Gambar I-1 6 (Enam) Bar pertama dari Partitur Mars ITB  
(Dokumentasi Pribadi)

Masing-masing dari golongan suara memiliki tugas menyanyikan nada yang berbeda-beda pula. Sopran biasa bertugas menyanyikan nada yang relatif tinggi bagi wanita, alto biasa bertugas menyanyikan nada yang relatif rendah bagi wanita, tenor biasa bertugas menyanyikan nada yang relatif tinggi bagi pria, dan bass biasa bertugas menyanyikan nada yang relatif rendah bagi pria.

Harmoni yang dihasilkan oleh kolaborasi berbagai golongan suara menciptakan perpaduan yang indah dan memanjakan sukma. Contohnya dalam Gambar 1-1, terdapat perbedaan nada yang dinyanyikan oleh sopran, alto, tenor, dan bass dalam waktu yang sama. Namun tentunya, terdapat keterbatasan seorang individu untuk masuk ke dalam suatu golongan suara. Seorang penyanyi profesional biasa mampu menyanyikan nada 2 oktaf, dan ini berarti hampir tidak mungkin seorang penyanyi untuk dapat masuk ke lebih dari 1 golongan suara walaupun tentunya dengan latihan yang tepat, dapat meningkatkan jangkauan nada seorang penyanyi.

Pemilihan golongan suara yang tidak tepat dapat sangat merugikan dari sisi seni maupun kesehatan pita suara dari sang penyanyi. Ketika seorang penyanyi memaksakan nada yang terlalu tinggi ataupun terlalu rendah untuk *range* yang

dijangkaunya, hal itu akan sangat membebani pita suara, dan dapat menyebabkan cacat permanen. Maka sangat penting menentukan golongan suara dengan tepat tergantung karakteristik dari suara penyanyi.

Karakteristik suara-pun dapat dikuantifikasi dalam bentuk representasi nilai (dB) dalam suatu frekuensi (hz) sedemikian rupa sehingga dapat dilakukan analisis numerik terhadap data suatu suara. Dengan memanfaatkan *Artificial Neural Network* (ANN) untuk meninjau faktor-faktor dalam suara, dapat dilakukan klasifikasi golongan suara dalam paduan suara. Hal ini sudah terbukti berdasarkan penelitian jurnal berjudul “*Deep Learning Approach for Singer Voice Classification of Vietnamese Popular Music*” menggunakan algoritma *Recurrent Neural Network* (RNN) yang mana menghasilkan *mean precision* senilai 85.4% (Van, T.P., dkk, 2019) serta dalam jurnal berjudul “*Human Vocal Type Classification using MFCC and Convolutional Neural Network.*” yang menggunakan model *Convolutional Neural Network*, berhasil mendapatkan akurasi sebesar 91.14% (Pratama, K.B., dkk, 2021).

Namun baik RNN maupun CNN memiliki kelemahan, yakni untuk RNN yaitu berbasis sekuensial, sehingga hubungan antar elemen yang berada pada satu *timeseries* yang sama tidak diperhitungkan. Sementara untuk CNN yaitu ketiadaan elemen *timeseries*. Hal ini menyebabkan terdapat beberapa aspek yang tidak dapat ditinjau oleh model, seperti kestabilan dan kepresisian nada. Terdapat suatu model yang acapkali digunakan untuk meninjau data yang memiliki aspek *timeseries* namun tetap memperhatikan hubungan antar fitur yang berada dalam satu sequence yang sama, yaitu *Convolutional Recurrent Neural Network*. Model *Convolutional Recurrent Neural Network* (CRNN) menggabungkan kelebihan dari kedua arsitektur, yaitu kemampuan *RNN* untuk menggali informasi sekuensial dan kemampuan *CNN* untuk menggali fitur spasial. Hal ini dibuktikan dalam jurnal berjudul “Implementasi Model Rekognisi Suara Menggunakan Metode *Convolutional Recurrent Neural Network* (CRNN)” (Octavya, 2021). Oleh sebab itu, diajukan model *Convolutional Recurrent Neural Network* untuk melakukan klasifikasi golongan suara dalam paduan suara dengan lebih akurat dan efisien.

## **I.2 Rumusan Masalah**

Berdasarkan latar belakang yang sudah dijelaskan di atas, yaitu kebutuhan akan teknik klasifikasi golongan suara dalam paduan suara, maka masalah utama yang difokuskan dalam penelitian ini adalah penentuan model klasifikasi golongan suara dalam paduan suara. Adapun rumusan masalah dalam tugas akhir ini adalah sebagai berikut :

1. Bagaimana cara membangun model pengklasifikasian golongan suara dalam paduan suara menggunakan *convolutional recurrent neural network* dan dengan menggunakan *convolutional neural network* serta *recurrent neural network*?
2. Bagaimana perbandingan hasil kinerja *convolutional recurrent neural network* dengan *convolutional neural network* serta *recurrent neural network* dalam konteks pengklasifikasian golongan suara dalam paduan suara tersebut?

## **I.3 Tujuan**

Berdasarkan rumusan masalah yang telah dijabarkan pada sub-bab I.2, maka didefinisikan beberapa tujuan Tugas Akhir sebagai berikut :

1. Membangun model pengklasifikasian golongan suara dalam paduan suara menggunakan *convolutional recurrent neural network*.
2. Membangun model pengklasifikasian golongan suara dalam paduan suara pembanding (*baseline*) menggunakan *convolutional neural network* serta *recurrent neural network*.
3. Membandingkan hasil kinerja *convolutional recurrent neural network* dengan *convolutional neural network* serta *recurrent neural network* dalam konteks pengklasifikasian golongan suara dalam paduan suara.

## **I.4 Batasan Masalah**

Berdasarkan rumusan masalah yang telah dijabarkan pada sub-bab I.2, batasan masalah yang diambil dalam pelaksanaan Tugas Akhir sebagai berikut:



1. Penyanyi yang dijadikan data latih dan data uji merupakan penyanyi yang dapat menyanyikan nada dengan tepat (tidak *tone deaf*).
2. *Labeling* terhadap *dataset* bersifat subjektif berdasarkan pengamatan dan *reasoning* pribadi tim pelatihan PSM-ITB.

## **I.5 Metodologi**

Tahapan-tahapan yang dipilih untuk menyelesaikan masalah dalam pengklasifikasian golongan suara dalam paduan suara adalah sebagai berikut :

1. *Planning*

Tahapan awal dalam penelitian tentang pengklasifikasian golongan suara dalam paduan suara akan dicari dan dianalisis solusi-solusi yang dapat ditawarkan. Beberapa solusi yang telah dikumpulkan kemudian dipilih yang bersesuaian dengan tujuan penelitian.

2. *Preparation*

Tahapan selanjutnya adalah pengumpulan dan pengolahan data latih dan data uji. Data-data ini diambil dari sumber internal anggota PSM-ITB yang telah dilabelkan golongan suaranya oleh tim pelatihan PSM-ITB. Keragaman label golongan suara yang diambil diharapkan mampu menghasilkan model yang dapat mengenali berbagai karakter suara.

3. *Designing*

*Designing* merupakan tahap untuk mendesain tahap-tahap yang harus dilakukan dalam melakukan *training* yang digunakan. Desain ini termasuk namun tidak terbatas pada pemilihan metode *training* serta penentuan arsitektur mesin pembelajaran.

4. *Training*

*Training* merupakan tahap yang dilakukan untuk pelatihan model. Pada pelatihan model ini, dilakukan perubahan penggunaan data untuk melakukan *validation* serta *hyperparameter tuning*.

## 5. *Analysis*

Model terbaik yang telah ditemukan diuji pada setiap skenario yang telah ditentukan dengan data uji. Selanjutnya hasil dari eksperimen tersebut dianalisis.

## **I.6 Sistematika Pembahasan**

Bab I merupakan bab Pendahuluan yang berisikan segala sesuatu mengenai alasan dibutuhkannya klasifikasi golongan suara dalam paduan suara menggunakan CRNN. Dalam bab ini pula terdapat rumusan masalah, tujuan, batasan masalah, hingga metodologi dalam penelitian ini.

Bab II merupakan bab Studi Literatur yang berisikan informasi literatur mengenai komponen-komponen pendukung dalam pembangunan solusi. Komponen pertama adalah karakteristik suara manusia, komponen kedua adalah data audio, komponen ketiga adalah metode klasifikasi suara manusia, komponen keempat adalah *convolutional recurrent neural network*, komponen kelima adalah hasil penelitian terkait, dan komponen terakhir adalah metrik evaluasi.

Bab III merupakan bab Analisis dan Rancangan klasifikasi golongan suara dalam paduan suara menggunakan CRNN. Pada bab ini dibahas secara mendalam mengenai permasalahan yang dihadapi pada model klasifikasi golongan suara yang sudah ada sekarang. Kemudian dibahas pula perihal bentuk dari solusi yang dibangun dalam penelitian serta cara mencapainya dengan lebih merinci.

Bab IV merupakan bab Evaluasi hasil pembelajaran dari model klasifikasi golongan suara dalam paduan suara menggunakan CRNN yang telah dibangun. Bab ini membahas mengenai bagaimana kinerja dari model klasifikasi golongan suara dalam paduan suara menggunakan CRNN yang telah dibangun. Kinerja ini dilihat dari hasil pengujian secara objektif, yaitu berdasarkan metrik evaluasi berupa *accuracy*, *recall*, *precision*, serta *f1-score*. Lalu terakhir analisis terhadap model yang telah dibuat dilakukan berdasarkan nilai metrik evaluasi yang didapat.

Bab V merupakan bab terakhir, yaitu bab Kesimpulan dan Saran. Bab ini berisikan kesimpulan mengenai apa yang bisa dicapai dari solusi yang dibangun. Selain itu, dibahas pula mengenai hal yang bisa ditingkatkan lagi di penelitian-penelitian selanjutnya. Bab ini merupakan bab penutup rangkaian Tugas Akhir.

## **BAB II**

### **STUDI LITERATUR**

Pada bab Studi Literatur, akan dijelaskan perihal berbagai studi literatur yang telah dilakukan dari berbagai sumber yang terkait terhadap apa yang diteliti dalam tugas akhir ini. Studi yang dibahas adalah mengenai pengetahuan musik serta metode klasifikasi golongan suara pada paduan suara.

#### **II.1 Karakteristik Suara Manusia**

Suara manusia dihasilkan oleh getaran dari pita suara manusia (Pratama, K. B., dkk, 2021). Hal yang sama juga berlaku terhadap suara nyanyian. Suara nyanyian dihasilkan oleh getaran pita suara manusia yang memiliki ritme dan nada tertentu untuk menghasilkan alunan melodi yang indah.

##### **II.1.1 Golongan Suara dalam Paduan Suara**

Paduan suara merupakan gabungan beberapa penyanyi yang menyatukan berbagai ragam jenis suara membentuk alunan melodi yang secara paralel akan membuat harmoni yang indah. Komposisi peserta paduan suara bergantung pada jenis, ukuran, serta kultur dari paduan suara itu sendiri. Ada beberapa paduan suara homogen (pria atau wanita saja) ataupun paduan suara heterogen (pria dan wanita). Karena adanya berbagai jenis paduan suara yang berbeda-beda, maka dibuat standarisasi golongan suara yang umum digunakan, yakni sopran dan alto untuk wanita, serta tenor dan bass untuk pria. Masing-masing golongan suara memiliki batasan *range vocal* rata-rata yang menandai seberapa tinggi/rendah suatu golongan suara mampu membunyikan nada. Hal ini juga dijadikan acuan dalam membuat partitur paduan suara sedemikian rupa sehingga partitur dalam paduan suara ini sendiri tidak membebani penyanyi dari suatu golongan suara tertentu.

Tabel II-1. *Vocal Range* untuk Tiap Golongan Suara berdasarkan The Concise Oxford Dictionary of Music (2007)

Golongan Suara	<i>Vocal Range</i> (Nada)	<i>Vocal Range</i> (Frekuensi)
Sopran	C4 – C6	261.63 – 1046.50
Alto	F3 – F5	174.61 – 698.46
Tenor	C3 – C5	130.81 – 523.25
Bass	E2 – E4	82.41 – 329.63

Berdasarkan Tabel II-1, terlihat bahwa terdapat beberapa nada yang *overlap*, misalnya nada terendah sopran masih berada di bawah nada tertinggi bass, hal ini menyebabkan terkadang terjadi kesalahan penempatan golongan suara dalam paduan suara.

### II.1.2 *Vocal overexertion*

*Vocal overexertion* adalah kondisi yang terjadi ketika seorang individu menggunakan suaranya secara berlebihan atau tidak sehat. Hal ini dalam jangka pendek dapat menyebabkan berbagai masalah seperti suara serak, sakit tenggorokan, hingga efek jangka panjang seperti polip dan kista (Johns Hopkins Medicine, 2023).

Pengalokasian golongan suara pada dasarnya harus dilakukan berdasarkan karakteristik serta jangkauan nada terkini guna menghindari kemungkinan terjadinya *vocal overexertion*. Namun pada kenyataannya, banyak praktik dalam paduan suara yang justru menentukan golongan suara berdasarkan kebutuhan golongan suara dan stereotip gender (Fisher, 2020).

## II.2 Data Audio

Data audio, khususnya dalam konteks bernyanyi, memiliki banyak fitur yang dapat dianalisis. fitur-fitur ini melibatkan berbagai aspek seperti *pitch* (nada), durasi, dan *timbre* (warna suara), serta berbagai fitur teknis lainnya yang bisa diekstraksi dari data audio tersebut. Berbeda dari standar nada normal yang umumnya diketahui, suara manusia tidak terdiri dari hanya satu gelombang. Bila seorang manusia

membunyikan nada A3, pada kenyataannya nada A3 hanyalah frekuensi dasar (*pitch / fundamental frequency /  $f_0$* ).

Setiap manusia memiliki karakteristik suara masing-masing yang menyebabkan warna suara (*timbre*) yang berbeda-beda pula. Dalam *music information retrieval* (MIR), salah satu aspek yang memengaruhi *timbre* adalah *harmonic series* ( $f_1, f_2, f_3, \dots, f_n$ ) dengan rasio tiap frekuensi adalah  $N/N-1$ . Pada kenyataannya, teori *harmonic series* tidak selalu sesuai dengan kenyataan. Ada konsep yang biasa disebut sebagai *inharmonic* dimana ada ketidaksesuaian antara frekuensi teori dengan frekuensi yang didapatkan di dunia nyata. Banyak aspek yang menyebabkan *inharmonic*, misalnya faktor eksternal yang memengaruhi gelombang suara. Contohnya bila  $f_0 = 440$  hz (A4), maka  $f_1$  secara teori adalah 880 hz (A5). Namun karena *inharmonic*, menyebabkan  $f_1$  menjadi 888 hz (Cano, E., dkk, 2018).

### II.2.1 Sumber Data Audio

Data audio yang digunakan berupa nyanyian nada C3 dan C4 bagi pria, serta C4 dan C5 bagi wanita. Durasi dari sumber data ini minimal 1.75 detik untuk tiap nadanya, namun disarankan lebih panjang sedemikian rupa dapat dilakukan *cutting* terhadap data tersebut.

Rekaman dibuat dalam format .WAV. WAV merupakan singkatan dari *waveform audio file format*. WAV dipilih sebagai format perekaman karena WAV menyimpan data audio dalam kualitas yang tidak dikompresi, yang berarti data audio tersimpan dalam kualitas aslinya, tanpa kehilangan data apapun. *Sample rate* yang digunakan dalam data audio ini adalah 22050 hz.

Data audio bersumber berdasarkan *crowdsourced* yang dilakukan kepada anggota Paduan Suara Mahasiswa ITB dengan perangkat perekaman pribadi responden. Proses pengumpulan data audio ini ditujukan khususnya kepada anggota Paduan Suara Mahasiswa ITB karena pelabelan data yang secara alami dilakukan di tiap penerimaan anggota Paduan Suara Mahasiswa ITB. Pengumpulan data dengan

metode *crowdsourced* ini sudah sering dilakukan dalam penelitian lain, terutama dalam anotasi audio (Pavlichenko, N, dkk, 2021).

Variabilitas dalam data audio ini berupa golongan suara dari responden. Seperti yang dijelaskan pada bagian II.1.1, terdapat 4 label yang menjadi identitas dari responden, yakni sopran, alto, tenor, dan bass. Terdapat pula variabilitas yang merupakan variabilitas non-teknikal seperti usia, jenis kelamin, serta tingkat kebisingan latar belakang.

## **II.2.2 Fitur Audio**

Data audio harus terlebih dahulu diubah dalam bentuk yang dapat dianalisis. Fitur yang akan digunakan dalam penelitian ini adalah *timbre* dan kestabilan suara, oleh karena itu, akan diambil fitur *mel-frequency cepstral coefficient*. *Best practice* yang sering digunakan adalah dengan mengubahnya terlebih dahulu ke bentuk *time-frequency* (TF). Kemudian dalam bentuk ini diubah lagi menjadi satuan spektogram. Data yang sudah dalam bentuk spektogram ini nantinya dapat diolah lebih lanjut dalam bentuk *mel-frequency cepstral coefficient* (MFCC).

### **II.2.2.1 Transformasi Time-Frequency (TF)**

Pengubahan data audio menjadi bentuk *time-frequency* dapat dilakukan dengan metode *fourier transform*. *Fourier transform* mengubah data audio menjadi beberapa data fungsi gelombang sinusoidal. Terdapat beberapa algoritma yang dapat digunakan untuk melakukan *fourier transform*. Salah satunya adalah algoritma *Short-Time Fourier Transform*. *Short-Time Fourier Transform* digunakan untuk memisahkan data audio ke dalam beberapa *layer* dengan 2 spektrum yaitu *real* dan *imaginary* (Rafii, Zafar, 2014).

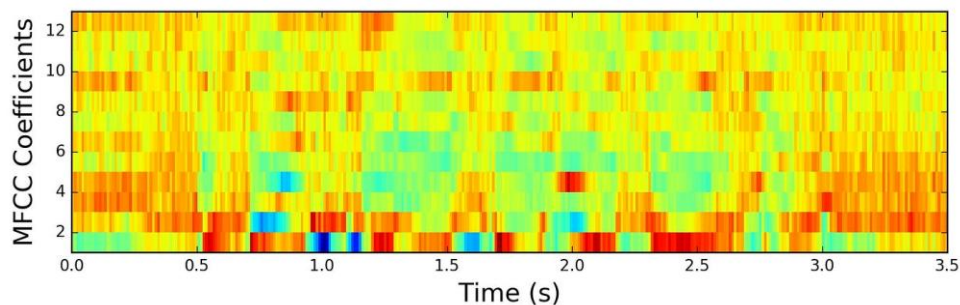
*Layer* pada indeks ke-0 disebut dengan *DC Component* yang mana nilainya selalu bersifat *real*. *Layer* pada indeks ke-1 hingga indeks ke-(N/2) berisikan nilai kompleks ( $a + j * b$ ). Sedangkan *layer* indeks ke-(N/2) hingga indeks ke-N-1 berisikan suatu nilai kompleks cerminan ( $a + j * b$ ). Bilamana N *layer* berjumlah genap, maka frekuensi pada *pivot* (N/2) bernilai selalu *real*.

### II.2.2.2 Besaran Spektrogram

Besaran spektrogram memiliki suatu nilai bilangan kompleks yakni  $(a + b * j)$  (Rafii, Zafar, 2014). Seluruh frekuensi dari nilai  $N$  merupakan bilangan *real* yang positif. Hasil kalkulasi pada nilai kompleks ini memiliki nilai yang berulang (*redundant*) pada frekuensi indeks ke- $(N/2)$  hingga indeks ke  $N-1$ . Sedangkan pada indeks lainnya merupakan nilai unik sesuai hasil kalkulasi nilai kompleks.

### II.2.2.3 Mel-Frequency Cepstral Coefficient (MFCC)

*Mel-Frequency Cepstral Coefficient* (MFCC) merupakan fitur yang sering digunakan dalam tugas-tugas *speech processing*. MFCC bekerja dengan mensimulasikan cara kerja telinga manusia dengan menangkap fitur-fitur yang berisikan informasi dari suara. Informasi ini termasuk namun tidak terbatas pada level energi dari *frequency band*, *spectral envelope*, serta *spectral shape*. MFCC dikenal mampu menangkap karakter (*timbre*) serta kestabilan dari suara manusia. (Van, T.P., dkk, 2019).



Gambar II-1 *Mel-Frequency Cepstral Coefficient*

Sumber: <https://medium.com/prathena/the-dummys-guide-to-mfcc-aceab2450fd>

MFCC biasanya terdiri dari 13 koefisien yang bertugas merepresentasikan tingkat energi total, frekuensi formant, serta amplitudo dari suara. Seperti pada gambar II-1, MFCC memiliki dimensi waktu (*time*) di dalamnya.



## II.3 Metode Klasifikasi Golongan Suara

Dalam pengklasifikasian golongan suara, akan dibangun 3 model berdasarkan pelatihan yang dilakukan menggunakan fitur yang telah diekstraksi. Dalam penelitian ini, model yang digunakan yakni *convolutional recurrent neural network* sebagai model utama, dan *convolutional neural network* serta *recurrent neural network* sebagai model pembanding.

Pembangunan model untuk pengklasifikasian golongan suara dalam paduan suara ini akan menggunakan bahasa pemrograman *python* dan memanfaatkan *library* yang ada. *Library* yang akan digunakan ini antara lain *librosa* untuk ekstraksi fitur serta *keras* untuk pemodelan.

### II.3.1 Ekstraksi Fitur

Ekstraksi fitur dilakukan menggunakan *librosa* dalam bahasa pemrograman *python*. Proses ekstraksi yang dilakukan adalah melakukan *short-time fourier transform* terhadap data audio dengan `librosa.stft`, kemudian ditranslasikan ke dalam bentuk *power* dengan `librosa.power_to_db`. Kemudian diambil bentuk *cepstrum*-nya menggunakan `librosa.feature.melspectrogram`. Terakhir bentuk tersebut ditranslasikan kembali ke bentuk *mfcc* dengan jumlah koefisien 13 menggunakan `librosa.feature.mfcc`.

```
import librosa

def extract_mfcc_features(audio, sample_rate, n_mfcc=13,
n_fft=2048, hop_length=512):
    stft = librosa.stft(audio, n_fft=n_fft,
hop_length=hop_length)
    power = librosa.power_to_db(np.abs(stft)**2)
    mel_spec = librosa.feature.melspectrogram(sr=sample_rate,
S=power)
    mfcc = librosa.feature.mfcc(S=librosa.power_to_db(mel_spec),
n_mfcc=n_mfcc)
    return mfcc

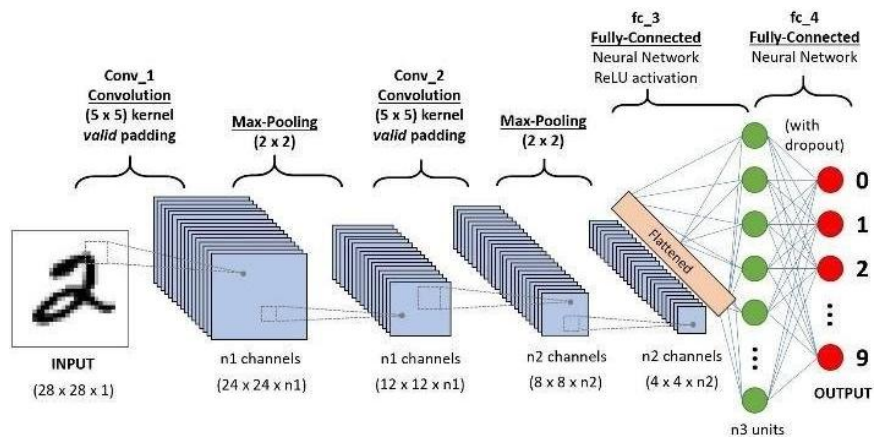
audio, sr = librosa.load(filename, sr=22050)
mfcc_features = extract_mfcc_features(audio, sr)
```

## II.3.2 Model Klasifikasi

Terdapat beberapa model klasifikasi yang dikerjakan dalam penelitian ini, yakni *baseline* berupa *convolutional neural network* dan *recurrent neural network*, serta model ajuan berupa *convolutional recurrent neural network*.

### II.3.2.1 Convolutional Neural Network

*Convolutional Neural Network* (CNN) merupakan salah satu jenis *Artificial Neural Network* (ANN) yang sering digunakan untuk menangkap hubungan antar fitur.



Gambar II-2 *Convolutional Neural Network*

Sumber: A Comprehensive Guide to Convolutional Neural Networks

Pada dasarnya seperti pada gambar II-2, CNN terdiri dari 4 jenis *layer*, yakni:

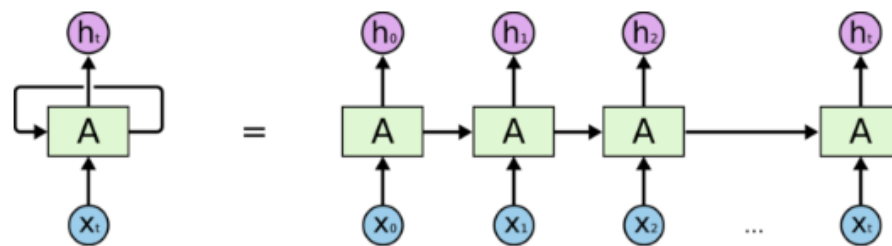
1. Lapisan *input*: Lapisan ini akan menerima data suara sebagai *input*, dan mengubahnya menjadi bentuk yang dapat diproses oleh jaringan saraf.
2. Lapisan *convolutional*: Lapisan ini akan menangkap pola-pola spasial dalam data suara dengan menggunakan filter-filter yang dapat mengidentifikasi fitur-fitur penting dalam data suara.
3. Lapisan *pooling*: Lapisan ini berfungsi untuk secara bertahap mengurangi ukuran spasial dari fitur yang diberikan lapisan sebelumnya. Jenis *pooling*

yang paling umum digunakan adalah *maxpooling*, yakni meneruskan hanya nilai terbesar dari kernel dengan ukuran tertentu.

4. Lapisan *output*: Lapisan ini akan menghasilkan prediksi golongan suara dalam paduan suara berdasarkan pola-pola yang telah ditangkap oleh lapisan *convolutional* sebelumnya.

### II.3.2.2 Recurrent Neural Network

*Recurrent Neural Network* merupakan salah satu jenis *Artificial Neural Network* (ANN) yang sering digunakan dalam menangkap keterhubungan urutan antar data, termasuk namun tidak terbatas pada *timeseries*. Berbeda dari jenis ANN lainnya, RNN mampu mengingat informasi dari langkah (*timeseries*) sebelumnya.



An unrolled recurrent neural network.

Gambar II-3 Recurrent Neural Network

Sumber: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>

Seperti pada gambar II-3, dapat dilihat bahwa RNN memiliki *loop* yang memungkinkan untuk meneruskan dan menerima informasi dari tahap/*timeseries* sebelum/sesudahnya.

### II.3.2.3 Convolutional Recurrent Neural Network

CRNN adalah sebuah metode pengklasifikasian suara yang menggabungkan teknik *convolutional neural network* (CNN) dan *recurrent neural network* (RNN). Metode ini pertama kali diperkenalkan oleh dua peneliti dari Google, dan telah digunakan

dalam beberapa konteks pengklasifikasian suara, seperti pengenalan suara manusia dan hewan.

CRNN menggabungkan kekuatan CNN untuk menangkap ciri-ciri lokal dari suara, dengan kekuatan RNN untuk menangkap keterkaitan temporer dalam suara. Ini memungkinkan metode ini untuk mengklasifikasikan suara dengan akurasi yang lebih tinggi daripada metode sebelumnya yang hanya menggunakan salah satu dari kedua jenis jaringan saraf ini.

Beberapa contoh penelitian yang menggunakan CRNN untuk mengklasifikasikan suara meliputi studi tentang pengenalan suara manusia yang diterbitkan pada tahun 2021 oleh Octavya pada jurnal “Implementasi Model Rekognisi Suara Menggunakan Metode Convolutional Recurrent Neural Network (CRNN)”, dan studi tentang pengenalan suara hewan yang diterbitkan pada tahun 2020 oleh N Singh dalam disertasi berjudul “*Classification of Animal Sound Using Convolutional Neural Network*”. Kedua studi ini menunjukkan bahwa CRNN mampu mengklasifikasikan suara dengan akurasi yang lebih tinggi daripada metode sebelumnya.

Secara umum, rencana arsitektur CRNN untuk pengklasifikasian golongan suara dalam paduan suara dapat terdiri dari beberapa komponen utama, seperti:

5. Lapisan *input*: Lapisan ini akan menerima data suara sebagai *input*, dan mengubahnya menjadi bentuk yang dapat diproses oleh jaringan saraf.
6. Lapisan *convolutional*: Lapisan ini akan menangkap pola-pola spasial dalam data suara dengan menggunakan filter-filter yang dapat mengidentifikasi fitur-fitur penting dalam data suara.
7. Lapisan *recurrent*: Lapisan ini akan menangkap pola-pola temporal dalam data suara dengan memproses data suara secara berulang seiring berjalannya waktu.
8. Lapisan *output*: Lapisan ini akan menghasilkan prediksi golongan suara dalam paduan suara berdasarkan pola-pola yang telah ditangkap oleh lapisan *convolutional* dan *recurrent* sebelumnya.

9. Lapisan penyempurnaan: Lapisan ini akan membantu meningkatkan akurasi model dengan melakukan pembelajaran ulang terhadap model CRNN dengan menggunakan data latih yang telah disediakan.

### II.3.3 Metrik Evaluasi

Metrik evaluasi yang digunakan dalam klasifikasi golongan suara adalah metrik evaluasi standar yang sering digunakan dalam klasifikasi. Dalam klasifikasi, metrik evaluasi yang sering digunakan adalah akurasi (*accuracy*), yaitu persentase kebenaran prediksi model dibandingkan dengan jumlah total data yang digunakan untuk evaluasi. Selain itu, metrik evaluasi lain yang sering digunakan dalam klasifikasi adalah presisi (*precision*) dan sensitivitas (*recall*). *Precision* mengukur seberapa sering model memprediksi suatu kelas dengan benar, sedangkan *recall* mengukur seberapa baik model dapat menemukan semua contoh dari suatu kelas. Kedua metrik *precision* dan *recall* bersamaan dapat membuat suatu metrik baru bernama *F1 Score* yang mana merupakan *harmonic mean* (rata-rata dengan pembobotan) dari *precision* dan *recall*. (Han. J, 2012).

Tabel II-2 *Confusion Matrix*

		<i>Predicted Value</i>	
		<b><i>Positive (P)</i></b>	<b><i>Negative (N)</i></b>
<i>Actual Value</i>	<b><i>Positive (P)</i></b>	<i>True-Positive (TP)</i>	<i>False-Negative (FN)</i>
	<b><i>Negative (N)</i></b>	<i>False-Postive (FP)</i>	<i>True-Negative (TN)</i>

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (\text{II.1})$$

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

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

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FP + FN} \quad (II.4)$$

Pada umumnya, metrik evaluasi ini digunakan bersama-sama untuk menilai seberapa baik suatu model dalam melakukan klasifikasi.

#### II.4 Hasil Penelitian Terkait

Pembuatan model klasifikasi golongan suara merupakan ilmu multidisiplin karena selain dari sisi *machine learning*, dibutuhkan juga sisi musikalitas dalam pengetahuan utama mengenai penggolongan suara (Wang, W, dkk, 2017). Sudah ada beberapa riset yang dijalankan mengenai klasifikasi suara dengan berbagai pendekatan dan berbagai metode yang mana hasilnya relatif baik (Kum, S, dkk, 2019). Pada beberapa jurnal, digunakan *Mel-Frequency Cepstral Coefficient* (MFCC) sebagai *feature extractor* dan menganalisis hasilnya menggunakan *Convolutional Neural Network* (CNN) sebagai *classifier*-nya. Penggunaan *Mel-Frequency Cepstral Coefficient* terbukti meningkatkan akurasi dari pengklasifikasian golongan suara hingga mencapai akurasi 91.14% (Pratama. K. B., dkk, 2021).

Sementara berdasarkan penelitian Van Pham, Toan, 2019, dengan menggunakan *Recurrent Neural Network* dalam penggolongan penyanyi di Vietnam, didapatkan bahwa penggunaan *Recurrent Neural Network* dengan analisis fitur *Mel-Frequency Cepstral Coefficient*, didapatkan hasil berupa nilai presisi 93%. Hal ini menandakan bahwa penggunaan *Recurrent Neural Network* dengan nilai presisi tertentu dapat digunakan dalam pengklasifikasian golongan suara *multi-label* dengan *target class* penyanyi itu sendiri.

Dari studi-studi tersebut, terbukti bahwasanya *Convolutional Neural Network* (CNN) dan *Recurrent Neural Network* (RNN) mampu menghasilkan hasil yang

cukup baik dalam klasifikasi data audio. Fokus utama dalam pemrosesan data audio adalah pengekstrakan fitur yang terkandung di dalam audio. Banyak metode yang digunakan untuk melakukan pengekstrakan ini, salah satunya yang paling sering digunakan adalah *Short-Time Fourier Transform* (STFT). Namun tidak menutup kemungkinan untuk menggunakan model seperti *convolutional neural network*, *artificial neural network*, ataupun *support vector machine* (Elbir, A, dkk, 2018).

Penggunaan *Convolutional Neural Network* (CNN) dan *Recurrent Neural Netowk* (RNN) sebagai basis dari pengklasifikasian suara sudah berkembang cukup pesat, kemudian terdapat pula modifikasi lebih lanjut seperti *Convolutional Recurrent Neural Network* (CRNN) yang terbukti menghasilkan hasil yang relatif baik dalam *voice recognition* bahkan melebihi tingkatan dari *convolutional neural network* serta *recurrent neural network* yang sudah cukup baik dalam pemrosesan data audio. Secara teoritis, *convolutional recurrent neural network* memiliki keunggulan dalam pemrosesan sinyal audio.

Lain halnya dengan penggunaan *Convolutional Recurrent Neural Network* pada Implementasi Model Rekognisi Suara Menggunakan Metode *Convolutional Recurrent Neural Network* (CRNN) yang dilakukan oleh Octavya pada tahun 2021, akurasi pada penggunaan *Convolutional Recurrent Neural Network* menghasilkan *training accuracy* sebesar 99.41% serta *testing accuracy* sebesar 99.05% dalam model rekognisi suara. Hal ini menandakan penggunaan *Convolutional Recurrent Neural Network* menghasilkan akurasi yang lebih baik dibandingkan model-model sebelumnya.

## **II.5 Kakas Pengembangan**

Dalam pengembangan model pembelajaran mesin pengklasifikasian golongan suara, terdapat kakas-kakas yang digunakan selama pengembangannya. Kakas-kakas ini dibagi berdasarkan tahap penggunaan kakas tersebut dalam pengembangan model pembelajaran mesin pengklasifikasian golongan suara.

### II.5.1 Kakas Praproses Data

Kakas yang digunakan dalam praproses data adalah *Audacity*. *Audacity* merupakan aplikasi pengeditan audio yang dikembangkan oleh *Audacity Team*. menggunakan bahasa C++. Fitur-fitur *Audacity* yang digunakan dalam praproses data antara lain fitur *cut* yang berfungsi untuk mengambil bagian tertentu dari audio, *padding* yang berfungsi menyamakan *timestamp* dari bagian data audio, serta fitur *compressor* yang berfungsi menormalisasikan *peak* audio.

### II.5.2 Kakas Ekstraksi Fitur

Kakas yang digunakan dalam ekstraksi fitur adalah *library librosa*. *Librosa* merupakan *library python* yang digunakan dalam analisis dan manipulasi audio. Dalam pengembangan model pembelajaran mesin pengklasifikasian golongan suara, fitur yang digunakan dari *librosa* adalah *librosa.stft* yang berfungsi untuk melakukan *short-time fourier transform*, *librosa.power\_to\_db* yang bertujuan mentranslasikan *power* ke dalam bentuk decibel. Kemudian *librosa.feature.melspectrogram* berfungsi untuk mengubah data menjadi bentuk *cepstrum*. Terakhir, *librosa.feature.mfcc* untuk mengubah data *cepstrum* menjadi bentuk *Mel-Frequency Cepstral Coefficient*.

### II.5.3 Kakas Pemodelan

Dalam melakukan pemodelan, kakas yang digunakan adalah *library keras*. *Keras* merupakan salah satu *library* dalam bidang *machine learning* yang paling sering digunakan dalam bahasa pemrograman *python*. *Keras* berisikan beberapa *layer* yang umum digunakan dalam *machine learning*, seperti Conv2D, MaxPooling2D, LSTM, dan lainnya. Oleh karena itu, *keras* digunakan dalam penelitian ini sebagai kakas pemodelan.



# **BAB III**

## **PENGEMBANGAN MODEL PEMBELAJARAN MESIN**

### **PENGKLASIFIKASIAN GOLONGAN SUARA**

Pada bab ini, akan dibahas mengenai analisis permasalahan pengklasifikasian golongan suara serta solusi pengembangan model pembelajaran mesin pengklasifikasian golongan suara dalam paduan suara beserta rancangannya.

#### **III.1 Analisis Persoalan**

Dalam pengklasifikasian suara pada paduan suara, terdapat beberapa aspek yang dapat memengaruhi golongan suara selain jenis kelamin dan *pitch*. Salah satunya adalah variabilitas suara atau lebih dikenal dengan kata *timbre*. Dalam pengklasifikasian golongan suara dalam paduan suara, *timbre* merupakan salah satu faktor yang patut dipertimbangkan. *Timbre* adalah sifat suara yang membedakannya dari suara lain, meskipun frekuensi dan intensitas yang sama. *Timbre* dapat diartikan sebagai "tampilan suara" atau "karakter suara" (Hanna & Deutsch, 2009).

Selain itu, stabilitas juga merupakan salah satu faktor yang perlu dipertimbangkan dalam pengklasifikasian golongan suara dalam paduan suara. Stabilitas merujuk pada kemampuan seseorang untuk menjaga nada yang tepat dalam suaranya, tanpa terlalu banyak variasi atau deviasi dari nada yang diinginkan. (Hanna & Deutsch, 2009).

Untuk aspek *timbre*, sebenarnya *Convolutional Neural Network* (CNN) sudah dapat digunakan untuk mendeteksinya. Seperti definisi pada bagian II.1, *timbre* dapat didefinisikan sebagai *harmonic series* dengan frekuensi tertentu yang bergetar secara sinkron dengan *pitch* ( $f_0$ ) karena salah satu karakteristik dari *Convolutional Neural Network* adalah kemampuannya dalam mendeteksi hubungan antar fitur

yang berdekatan (Pratama, K. B., dkk, 2021). Sementara untuk kestabilan suara bisa dideteksi dengan data masukan yang memiliki *timeseries* contohnya dengan menggunakan model *recurrent neural network* (RNN) dimana kestabilan suara dapat dilihat dari seberapa sempurna seseorang menahan suatu nada pada suatu interval waktu tertentu. Hal ini disebabkan karena salah satu karakteristik dari *recurrent neural network* (RNN) adalah kemampuannya mendeteksi keterurutan fitur dalam *timeseries* (Van, T. P., dkk, 2019).

Namun masing-masing arsitektur memiliki kelemahannya masing-masing. *Convolutional neural network* tidak dapat menangkap keterurutan fitur dalam *timeseries*, dan *recurrent neural network* tidak dapat menangkap hubungan antar fitur. Membuat model yang menggabungkan kemampuan menangkap keterurutan fitur dalam *timeseries* dan kemampuan menangkap hubungan antar fitur diharapkan memberikan hasil yang lebih baik dari model *convolutional neural network* dan *recurrent neural network*.

### **III.2 Analisis Solusi**

Metode pengklasifikasian golongan suara dalam paduan suara sebenarnya sudah masuk dalam tahap pencarian *state of art*. Sejauh ini sudah ada beberapa model yang dibuat dari berbagai arsitektur yang ada seperti *convolutional neural network*, *recurrent neural network*, hingga yang paling sederhana menggunakan *artificial neural network* yang hanya sebatas mencari *threshold pitch* terbaik. Namun sebenarnya model-model tersebut masih dapat dikembangkan lagi karena model-model tersebut belum sempurna.

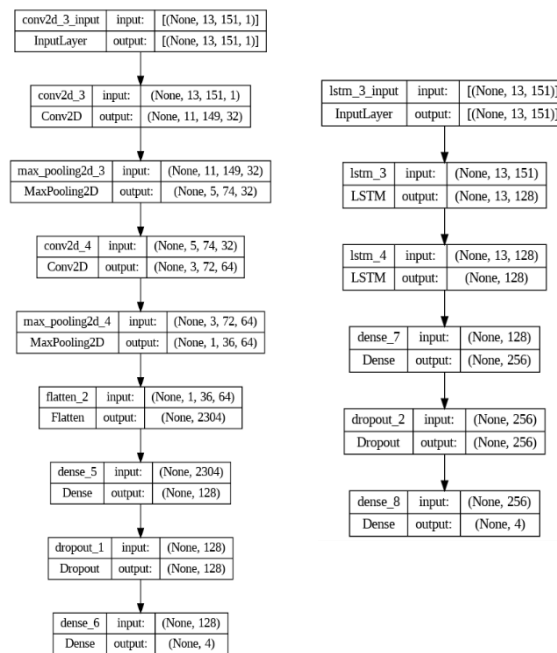
Salah satu model arsitektur yang dapat dipertimbangkan untuk mengatasi permasalahan yang ada adalah menggunakan arsitektur *Convolutional Recurrent Neural Network* (CRNN). CRNN merupakan arsitektur gabungan antara *convolutional neural network* dan *recurrent neural network* yang mana menggabungkan kelebihan dari kedua arsitektur tersebut yakni kemampuan menangkap hubungan antar fitur serta kemampuan menangkap keterurutan fitur dalam *timeseries*. Dengan menggabungkan CNN dan RNN menjadi arsitektur

CRNN, model yang dihasilkan diharapkan mampu mengenali *timbre* dan stabilitas suara dengan lebih baik dari model-model yang sudah ada sebelumnya.

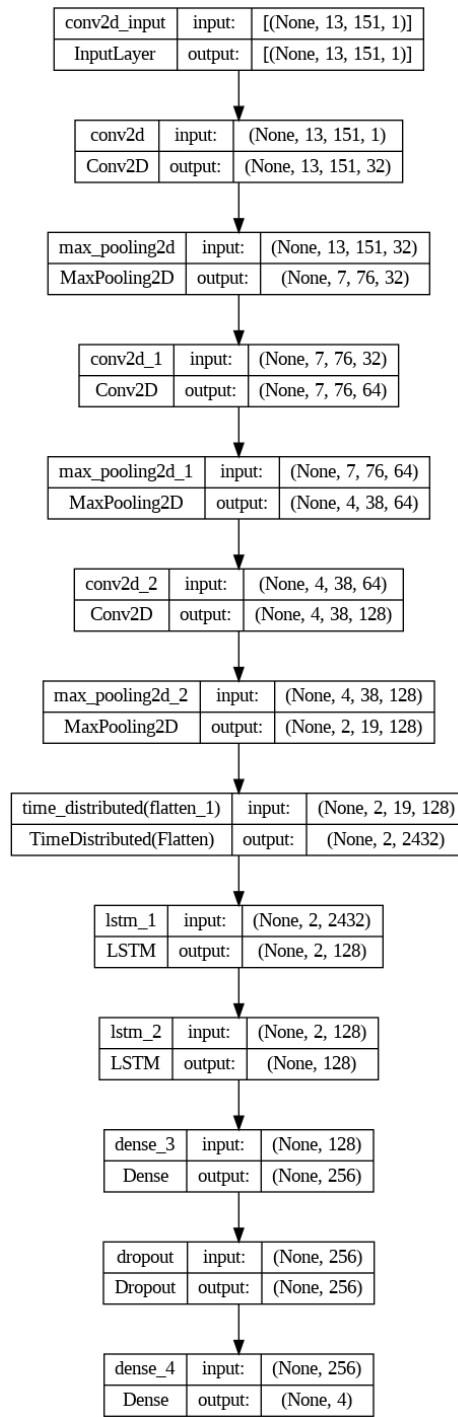
### III.3 Rancangan Solusi

Sama seperti solusi pembelajaran mesin lainnya, tahapan pembelajaran mesin dibagi menjadi beberapa tahap, seperti pengumpulan *dataset*, pengolahan data, pemodelan, desain eksperimen model, dan diakhiri dengan pengujian. Tahapan-tahapan ini digunakan untuk pembuatan model *convolutional recurrent neural network* (Gambar III-2) serta model *baseline*-nya yakni *convolutional neural network* dan *recurrent neural network* (Gambar III-1).

Berikut merupakan detail arsitektur dari CNN, RNN, serta CRNN menggunakan fitur *plot\_model* bawaan *tensorflow utils*:



Gambar III-1 Arsitektur *baseline* CNN (kiri) dan RNN (kanan)



Gambar III-2 Arsitektur *Convolutional Recurrent Neural Network*

### III.4 Pengumpulan *Dataset*

Sama seperti rancangan solusi pembelajaran mesin umumnya, tahap awal yang dilakukan dalam pengembangan model pembelajaran mesin pengklasifikasian golongan suara adalah tahap pengumpulan *dataset*. *Dataset* yang dikumpulkan bersumber anggota Paduan Suara Mahasiswa ITB (PSM-ITB). Dalam pembangkitan *dataset* ini, perekaman dilakukan dengan merekam nada C3 dan C4 bagi pria, serta C4 dan C5 bagi wanita. Pemilihan nada C, baik C3, C4, maupun C5 adalah karena berdasarkan tabel II-1, nada tersebut merupakan nada yang secara teoritis dapat dijangkau semua jenis golongan suara. Nada direkam selama paling singkat 1.75 detik untuk tiap nadanya. Perekaman tidak menggunakan ruang studio melainkan sebatas diberi instruksi untuk merekam menggunakan *device* masing-masing di ruangan yang senyap.

### III.5 Pengolahan Data

Data audio yang sudah dikumpulkan pertama diseragamkan melalui tahapan-tahapan *preprocessing*, yakni menggunakan aplikasi editor audio *Audacity*. *Preprocessing* yang dilakukan adalah melakukan *cutting* sedemikian rupa sehingga nada yang diambil memiliki panjang yang sama. Setelah itu bagian perpindahan nada dihapuskan sedemikian rupa sehingga bagian perpindahan nada nantinya tidak masuk ke dalam perhitungan. Selain itu diberikan *zero-padding* sedemikian rupa sehingga waktu mulai dan berakhir tiap nada berada di titik yang sama. Seluruh audio dikonversi ke dalam bentuk .wav dengan *sampling rate* senilai 22050 hz (22.05 khz).

Setelah melewati tahap *preprocessing*, dilakukan ekstraksi fitur *mel-frequent cepstral coefficients* dengan menggunakan *library python* bernama *librosa*. Tahap awal adalah melakukan *load* audio menggunakan fitur *load* dari *librosa*. Audio yang di-*load* oleh *librosa* akan berupa data numerik *waveform*. Data ini dimasukkan ke dalam fungsi *extract\_mfcc\_features* dimana fungsi ini berisikan serangkaian langkah untuk mengekstrak *mel-frequent cepstral coefficients* dengan *librosa*. Langkah pertama adalah dengan menghitung *short-time fourier transform* dengan

menggunakan fitur *stft* dari *librosa* dengan panjang jendela *fast forier transform* senilai 2048 dan panjang lompatan antar *frame* senilai 512. Kemudian berdasarkan hasil *stft*, dibuat spektrogram menggunakan fitur *librosa* yakni *power\_to\_db*. Spektrogram ini kemudian diubah ke bentuk *mel spectrogram* menggunakan fitur *librosa* yakni *melspectrogram*. Tahapan akhir adalah membuat representasi *Mel-Frequent Cepstral Coefficients* dengan fitur *librosa* yakni *mfcc* dengan *n\_mfcc* yang dipilih adalah 13 koefisien.

### III.6 Pemodelan Pengklasifikasian Golongan Suara

Hasil ekstraksi fitur MFCC yang sudah dilakukan sebelumnya, kemudian dimasukkan ke dalam algoritma pembelajaran mesin pengklasifikasian golongan suara dengan arsitektur *Convolutional Recurrent Neural Network* dengan arsitektur lengkap dapat dilihat pada gambar III-2 sementara untuk model *baseline Convolutional Neural Network* dan *recurrent Neural Network* menggunakan duplikasi arsitektur pada gambar III-1.

Pada tahap ini, fitur *mel-frequent cepstral coefficients* yang telah diekstrak sebelumnya dijadikan data latih menggunakan algoritma pembelajaran mesin CRNN. Layer pertama dalam CRNN berisikan *input layer* yang berfungsi menerima masukan berupa fitur MFCC dengan bentuk yang telah ditentukan berupa (*n\_mfcc*, jumlah frame, jumlah channel) atau dalam kasus ini (13,151,1).

Layer berikutnya merupakan layer *convolutional* dengan ukuran kernel (3,3). Layer ini menggunakan 32 unit dan seperti yang dijelaskan sebelumnya, tujuan utamanya adalah mendeteksi hubungan antar fitur lokal. Kemudian terdapat layer *maxpooling* yang bertujuan mengurangi dimensi dari fitur dengan mengambil nilai maksimum dari jendela dengan ukuran kernel (2,2). Layer *convolutional* dan *recurrent* ini diulang 2 kali lagi dengan ukuran unit yang meningkat yaitu pada pengulangan kedua terdapat 64 unit pada layer *convolutional*, dan pada pengulangan terakhir, terdapat 128 unit pada layer *convolutional*.

Setelahnya, diberikan layer *time\_distributed* dengan tujuan mengaplikasikan *flatten* pada setiap *time step* dari keluaran layer sebelumnya. Layer ini bertugas membuat hasil keluaran layer sebelumnya untuk dapat dimasukkan ke dalam *recurrent neural network*. Kemudian hasil dari layer ini diteruskan ke dalam layer *bidirectional LSTM*. Layer *bidirectional LSTM* ini menggunakan 128 unit dengan *dropout rate* sebesar 0.2. Seperti yang dijelaskan sebelumnya, layer *recurrent neural network* ini bertugas untuk mengenali pola di data *timeseries*. Layer setelahnya adalah layer *dense* atau dikenal juga dengan nama layer *fully connected* berfungsi untuk memetakan pengklasifikasian *bidirectional LSTM* ini ke dalam 256 unit. *Dropout rate* sebesar 0.5 diterapkan setelahnya untuk mencegah *overfitting*. Layer terakhir adalah layer *dense* lagi yang memetakan hasil layer sebelumnya ke dalam 4 kelas yang ada.

### III.7 Desain Eksperimen Model Pengklasifikasian Golongan Suara

Desain eksperimen model pengklasifikasian golongan suara dibuat sebagai acuan proses pelatihan model. Tujuan utama dari penggunaan desain eksperimen model ini adalah guna menciptakan model sebaik mungkin. Strategi eksperimen yang dilakukan adalah *one factor at a time* guna mengetahui pengaruh dari masing-masing parameter dalam pelatihan menggunakan *tuning* pada *hyperparameter*. *Hyperparameter* merupakan parameter yang diatur dan diisi sendiri sedemikian rupa sehingga memengaruhi kinerja model. *Hyperparameter* yang dipilih dalam eksperimen model ini adalah *epoch* dan juga *learning rate*.

### III.8 Pengujian

Guna mengukur performansi model yang dibangun, dilakukan evaluasi model pembelajaran mesin pengklasifikasian golongan suara. Evaluasi dilakukan dengan data uji menggunakan skema *random sampling*. Rasio yang digunakan adalah 80:20 untuk data latih dan data uji. Metrik evaluasi yang digunakan adalah *accuracy*, *precision*, *recall*, dan *f1-score*. Selain itu juga akan dilakukan evaluasi subjektif pada data uji.

## BAB IV

### EVALUASI MODEL PEMBELAJARAN MESIN

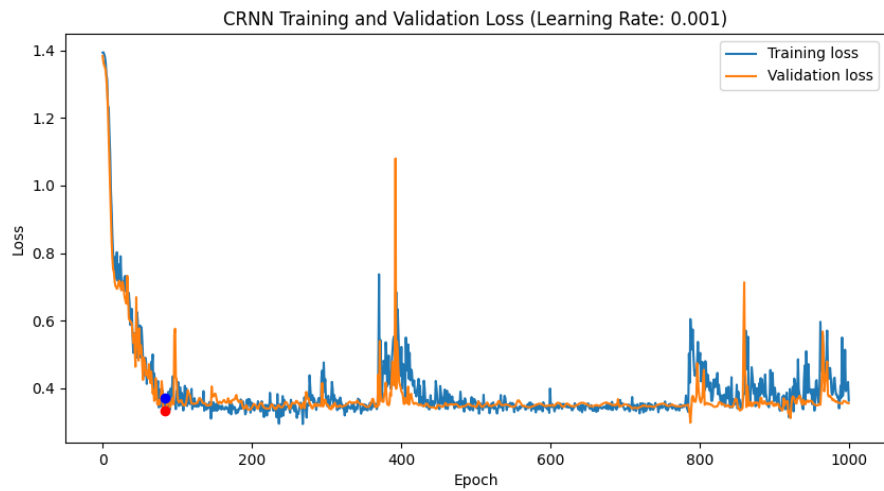
### PENGKLASIFIKASIAN GOLONGAN SUARA

Evaluasi model pembelajaran mesin pengklasifikasian golongan suara dibagi menjadi 3 bagian, yakni evaluasi tahap training dimana hasil *hyperparameter tuning* pada sebagian data latih ditampilkan, evaluasi tahap testing dimana hasil *training* dari data latih dengan menggunakan *parameter* yang sudah di-*tuning* sebelumnya ditampilkan, dan analisis kinerja dari model pembelajaran mesin pengklasifikasian golongan suara.

#### IV.1 Evaluasi Tahap Training

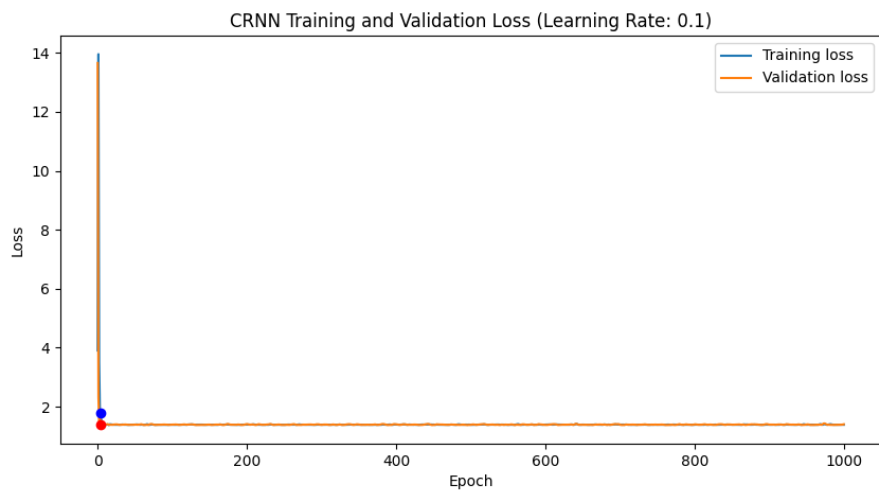
Dalam penelitian ini, dibuat sebuah algoritma sederhana yang melakukan eksperimen *hyperparameter tuning* dengan strategi *one factor at a time* pada tiap modelnya. *Tuning* yang dilakukan adalah menggunakan *early stop epoch* hingga 1000 dan *learning\_rate*=[0.1,0.001,0.0005,0.0001,0.00005,0.00001]. *Early stop* dilakukan pada kondisi *loss* dan *val\_loss* berdekatan, dan penurunan yang terjadi sudah tidak signifikan atau bahkan cenderung mengalami peningkatan pada *val\_loss*. Visualisasi *loss* dan *val\_loss* dapat dilihat di bawah ini:





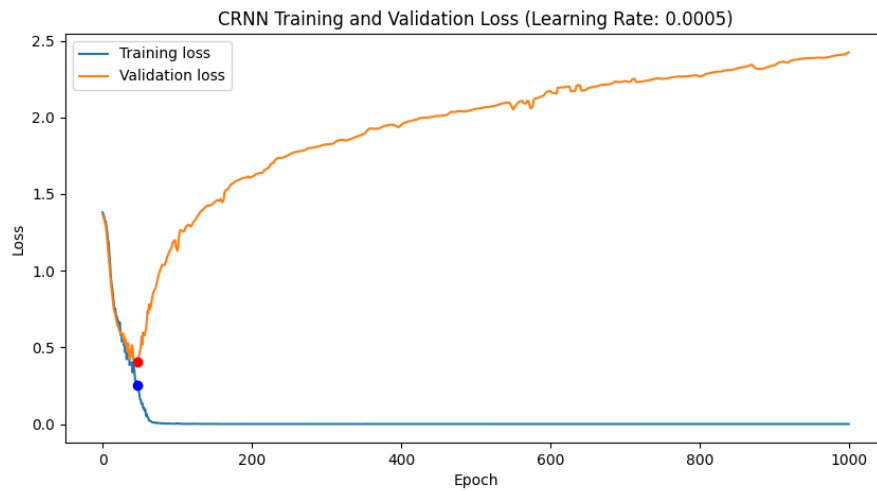
Gambar IV-1 Titik Seimbang model CRNN dengan LR 0.001

Pada model CRNN dengan *Learning Rate* 0.001, didapati titik *epoch* ke-84 ditandai pada gambar IV-1.



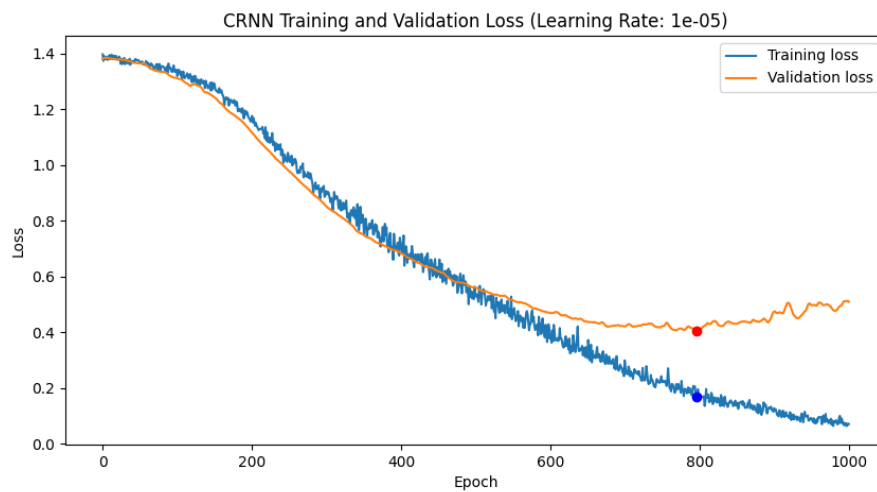
Gambar IV-2 Titik Seimbang Model CRNN dengan LR 0.1

Pada model CRNN dengan *Learning Rate* 0.1, didapati titik *epoch* ke-4 ditandai pada gambar IV-2.



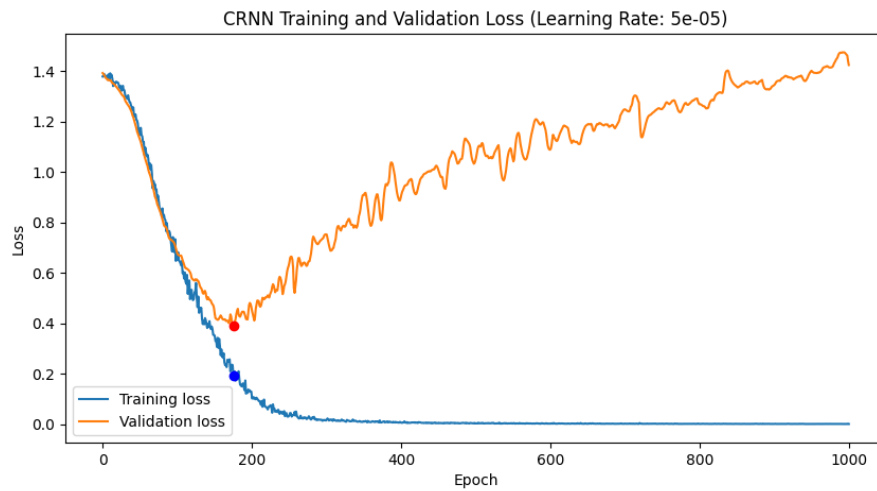
Gambar IV-3 Titik Seimbang Model CRNN dengan LR 0.0005

Pada model CRNN dengan *Learning Rate* 0.0005, didapati titik *epoch* ke-46 ditandai pada gambar IV-3.



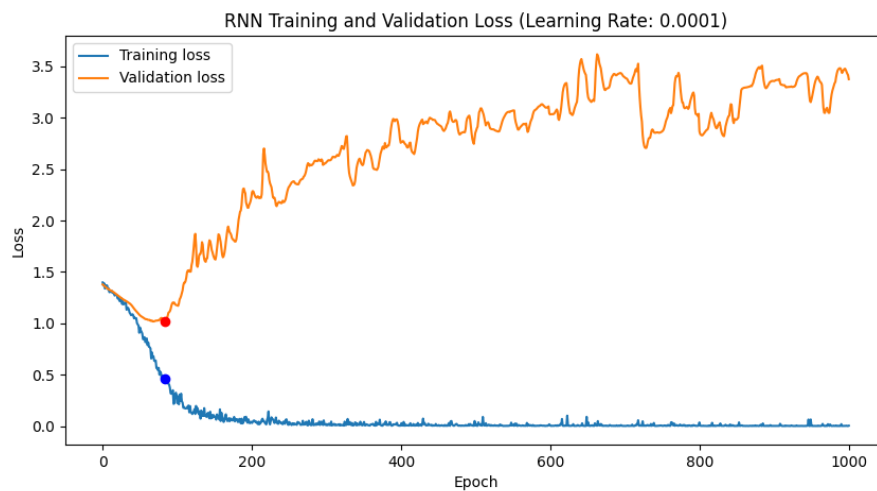
Gambar IV-4 Titik Seimbang Model CRNN dengan LR 0.00001

Pada model CRNN dengan *Learning Rate* 0.00001, didapati titik *epoch* ke-795 ditandai pada gambar IV-4.



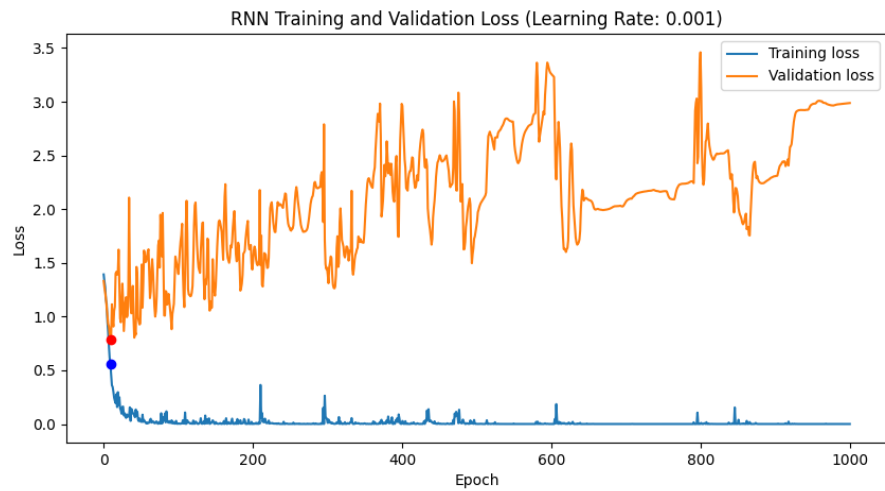
Gambar IV-5 Titik Seimbang Model CRNN dengan LR 0.00005

Pada model CRNN dengan *Learning Rate* 0.00005, didapati titik *epoch* ke-175 ditandai pada gambar IV-5.



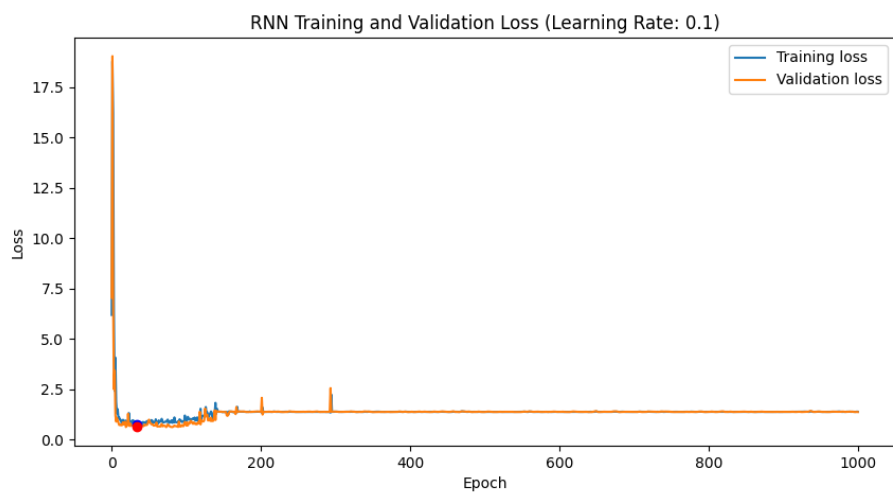
Gambar IV-6 Titik Seimbang Model RNN dengan LR 0.0001

Pada model RNN dengan *Learning Rate* 0.0001, didapati titik *epoch* ke-83 ditandai pada gambar IV-6.



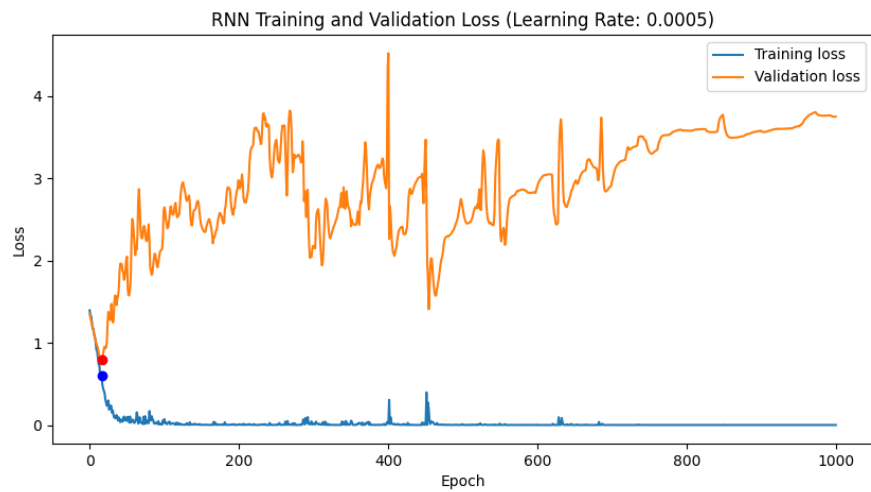
Gambar IV-7 Titik Seimbang Model RNN dengan LR 0.001

Pada model RNN dengan *Learning Rate* 0.001, didapati titik *epoch* ke-9 ditandai pada gambar IV-7.



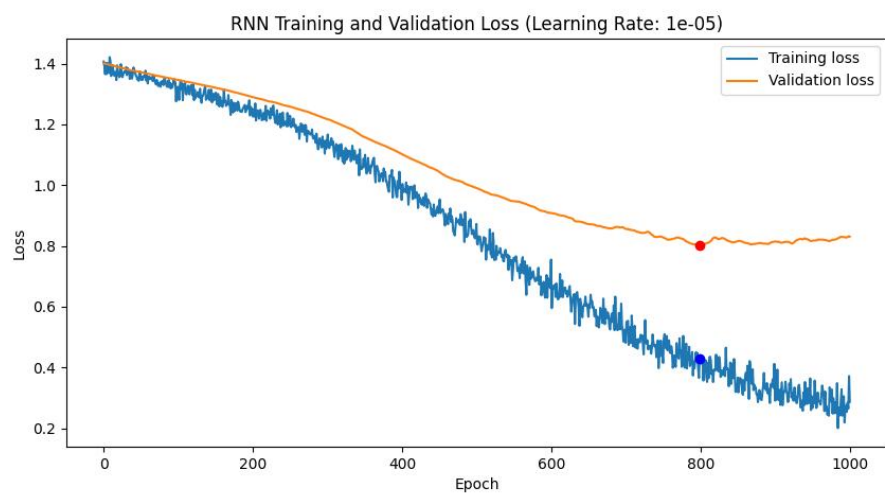
Gambar IV-8 Titik Seimbang Model RNN dengan LR 0.1

Pada model RNN dengan *Learning Rate* 0.1, didapati titik *epoch* ke-34 ditandai pada gambar IV-8.



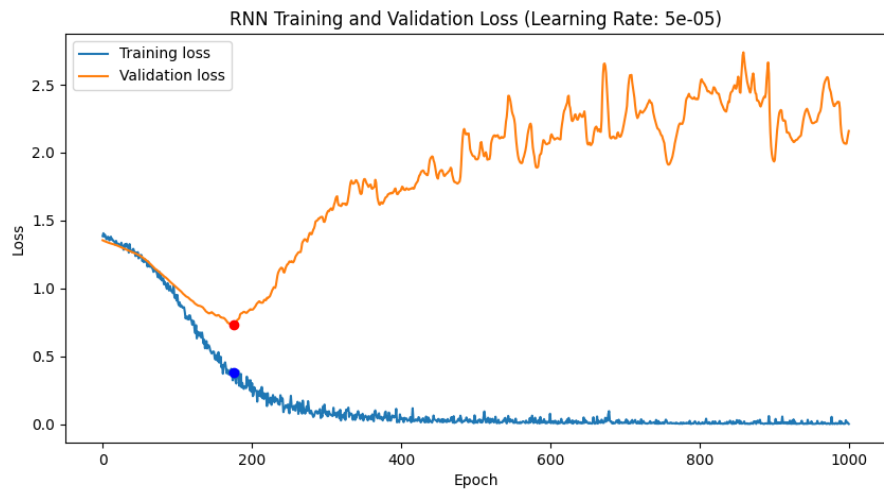
Gambar IV-9 Titik Seimbang Model RNN dengan LR 0.0005

Pada model RNN dengan *Learning Rate* 0.0005, didapati titik *epoch* ke-16 ditandai pada gambar IV-9.



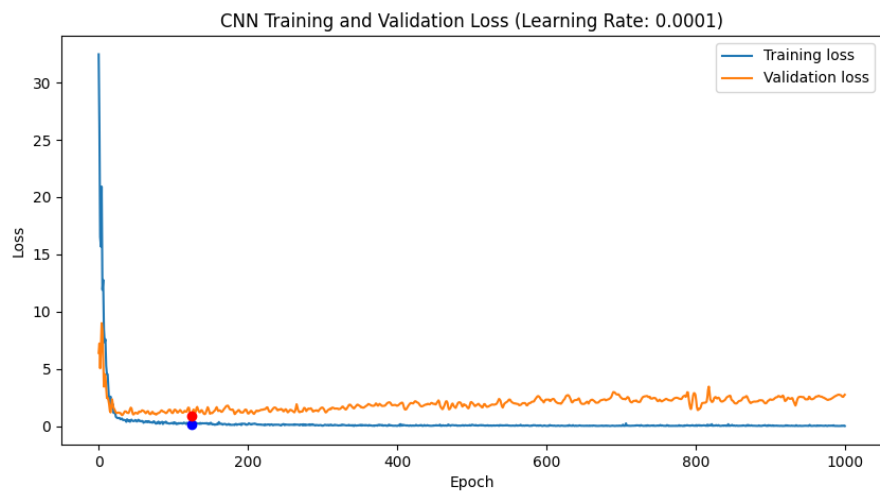
Gambar IV-10 Titik Seimbang Model RNN dengan LR 0.00001

Pada model RNN dengan *Learning Rate* 0.00001, didapati titik *epoch* ke-798 ditandai pada gambar IV-10.



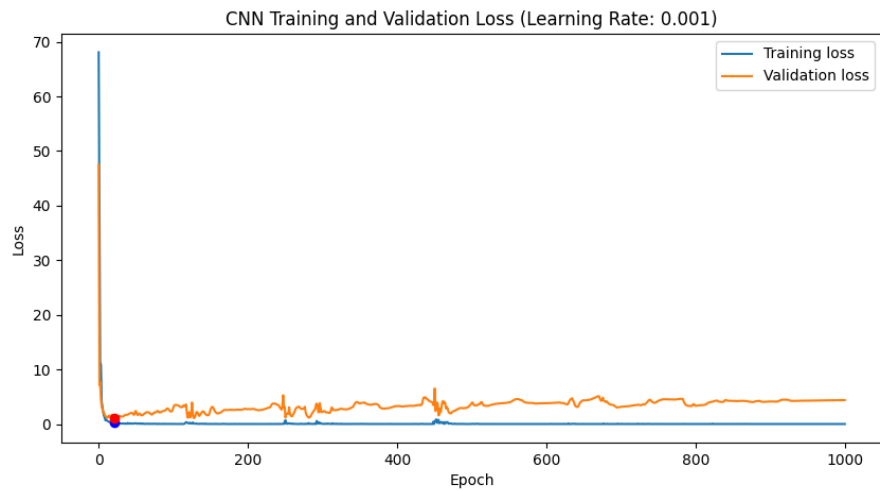
Gambar IV-11 Titik Seimbang Model RNN dengan LR 0.00005

Pada model RNN dengan *Learning Rate* 0.00005, didapati titik *epoch* ke-175 ditandai pada gambar IV-11.



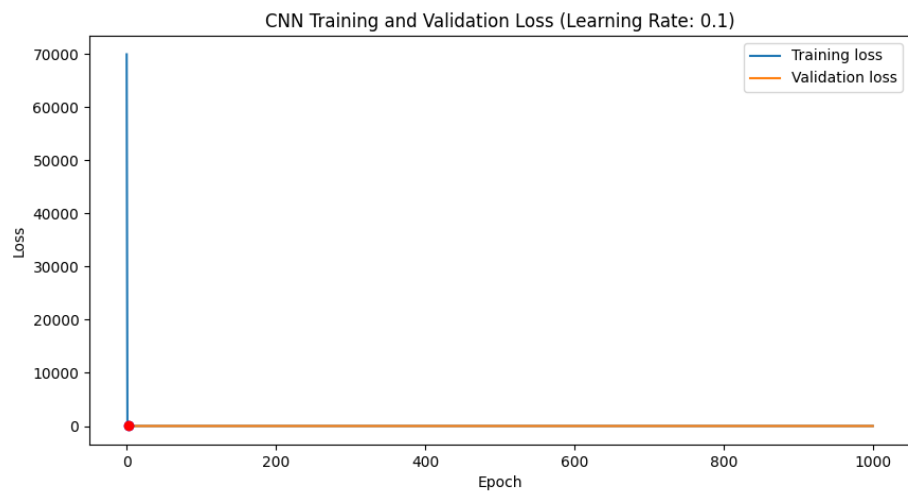
Gambar IV-12 Titik Seimbang Model CNN dengan LR 0.0001

Pada model CNN dengan *Learning Rate* 0.0001, didapati titik *epoch* ke-124 ditandai pada gambar IV-12.



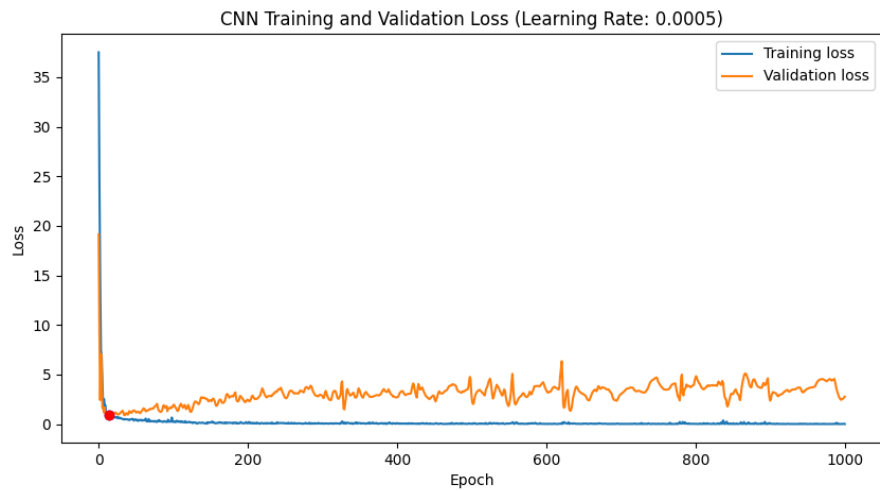
Gambar IV-13 Titik Seimbang Model CNN dengan LR 0.001

Pada model CNN dengan *Learning Rate* 0.001, didapati titik *epoch* ke-21 ditandai pada gambar IV-13.



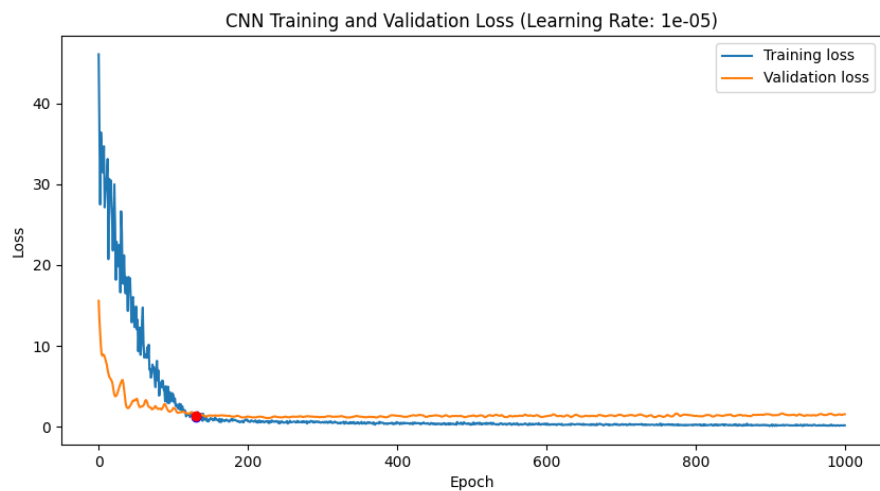
Gambar IV-14 Titik Seimbang Model CNN dengan LR 0.1

Pada model CNN dengan *Learning Rate* 0.1, didapati titik *epoch* ke-3 ditandai pada gambar IV-14.



Gambar IV-15 Titik Seimbang Model CNN dengan LR 0.0005

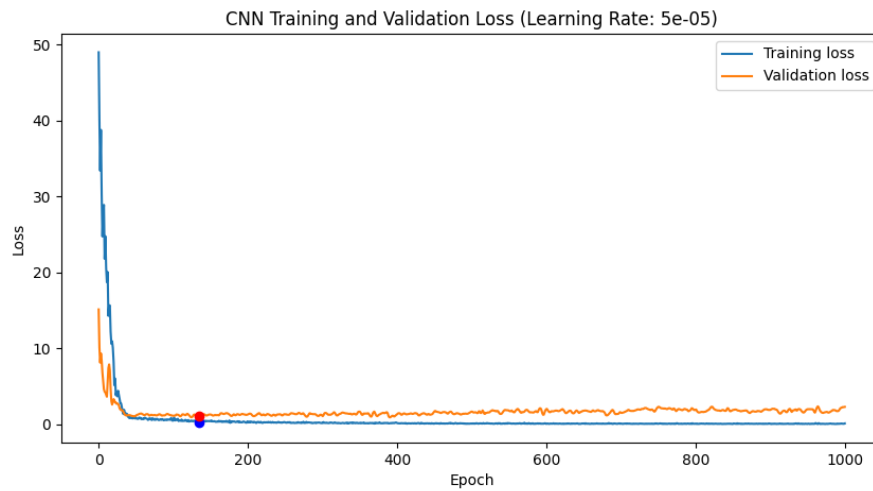
Pada model CNN dengan *Learning Rate* 0.0005, didapati titik *epoch* ke-14 ditandai pada gambar IV-15.



Gambar IV-16 Titik Seimbang Model CNN dengan LR 0.00001

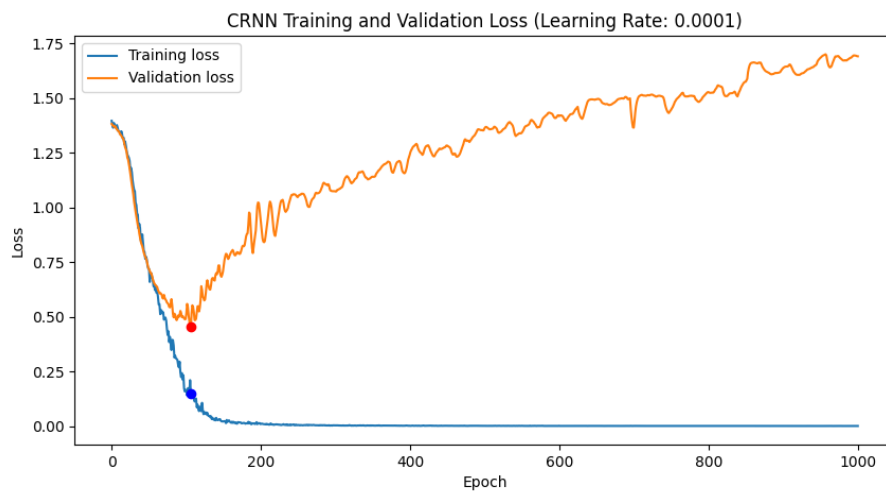
Pada model CNN dengan *Learning Rate* 0.00001, didapati titik *epoch* ke-130 ditandai pada gambar IV-16.





Gambar IV-17 Titik Seimbang Model CNN dengan LR 0.00005

Pada model CNN dengan *Learning Rate* 0.00005, didapati titik *epoch* ke-135 ditandai pada gambar IV-17.



Gambar IV-18 Titik Seimbang Model CRNN dengan LR 0.0001

Pada model CRNN dengan *Learning Rate* 0.0001, didapati titik *epoch* ke-106 ditandai pada gambar IV-18.

Berdasarkan visualisasi-visualisasi di atas, didapatkan *epoch* terbaik untuk masing-masing *learning rate* pada tabel IV-1 sebagai berikut:

Tabel IV-1 *Epoch* dengan *loss* optimal

	CRNN	RNN	CNN
0.1	4	34	3
0.001	84	9	21
0.0005	46	16	14
0.0001	106	83	124
0.00005	175	175	135
0.00001	795	798	130

Berdasarkan kombinasi *epoch* tersebut, dibandingkan data *accuracy* dan *val\_accuracy* sesuai tabel IV-2 berikut:

Tabel IV-2 *Accuracy* dan *val\_accuracy* untuk masing-masing *epoch* optimal

Model	Combination	accuracy	val_accuracy
CRNN	epoch_4_lr_0.1	0.2253521085	0.277777791
CRNN	epoch_84_lr_0.001	0.7042253613	0.777777791
CRNN	epoch_46_lr_0.0005	0.9295774698	0.6666666865
CRNN	epoch_106_lr_0.0001	0.9718309641	0.777777791
CRNN	epoch_175_lr_5e-05	0.985915482	0.777777791
<b>CRNN</b>	<b>epoch_795_lr_1e-05</b>	<b>0.985915482</b>	<b>0.8333333135</b>
RNN	epoch_34_lr_0.1	0.4788732529	0.6666666865
RNN	epoch_9_lr_0.001	0.7605633736	0.5
RNN	epoch_16_lr_0.0005	0.8309859037	0.4444444478
RNN	epoch_83_lr_0.0001	0.887323916	0.4444444478
<b>RNN</b>	<b>epoch_175_lr_5e-05</b>	<b>0.887323916</b>	<b>0.6666666865</b>
RNN	epoch_798_lr_1e-05	0.887323916	0.6666666865

CNN	epoch_3_lr_0.1	0.1971831024	0.277777791
CNN	epoch_21_lr_0.001	0.9014084339	0.722222209
CNN	epoch_14_lr_0.0005	0.549295783	0.555555582
<b>CNN</b>	<b>epoch_124_lr_0.0001</b>	<b>0.9718309641</b>	<b>0.722222209</b>
CNN	epoch_135_lr_5e-05	0.887323916	0.6666666865
CNN	epoch_130_lr_1e-05	0.6338028312	0.5

## IV.2 Evaluasi Objektif Tahap Testing

Berdasarkan evaluasi tahap *training*, dibuatlah model berdasarkan *hyperparameter tuning* terbaik (dicetak tebal pada Tabel IV-2) untuk model CRNN, CNN, dan RNN. Model dilatih dengan *training data* kemudian diuji menggunakan *testing data* berjumlah 23 data dengan pembagian 6 Sopran, 6 Alto, 5 Tenor, dan 6 Bass. Berikut merupakan *classification report* bawaan *sklearn* untuk CRNN (tabel IV-2), CNN (tabel IV-3) dan RNN (tabel IV-4).

Tabel IV-3 *Classification Report CRNN*

	<b>Precision</b>	<b>Recall</b>	<b>F1-score</b>
Sopran	0.67	1.00	0.80
Alto	1.00	0.50	0.67
Tenor	1.00	1.00	1.00
Bass	1.00	1.00	1.00
accuracy	0.87		

Tabel IV-4 *Classification Report CNN*

	<b>Precision</b>	<b>Recall</b>	<b>F1-score</b>
Sopran	0.62	0.83	0.71
Alto	0.75	0.50	0.60
Tenor	0.67	0.80	0.73
Bass	1.00	0.83	0.91
accuracy	0.74		

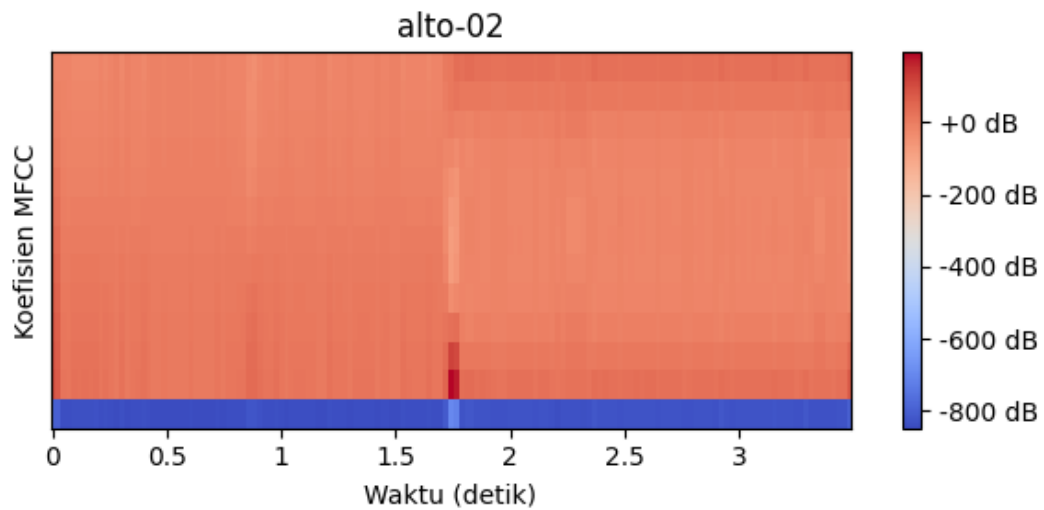
Tabel IV-5 *Classification Report RNN*

	<b>Precision</b>	<b>Recall</b>	<b>F1-score</b>
Sopran	0.67	0.67	0.67
Alto	0.60	0.50	0.55
Tenor	0.50	0.80	0.62
Bass	1.00	0.67	0.80
accuracy	0.65		

Berdasarkan hasil *classification report*, untuk tiap aspek, baik *precision*, *recall*, dan *f1-score* serta *accuracy* rata-rata dari seluruh kelas, didapati bahwa *convolutional recurrent neural network* menunjukkan hasil yang lebih baik daripada model *convolutional neural network* dan *recurrent neural network*.

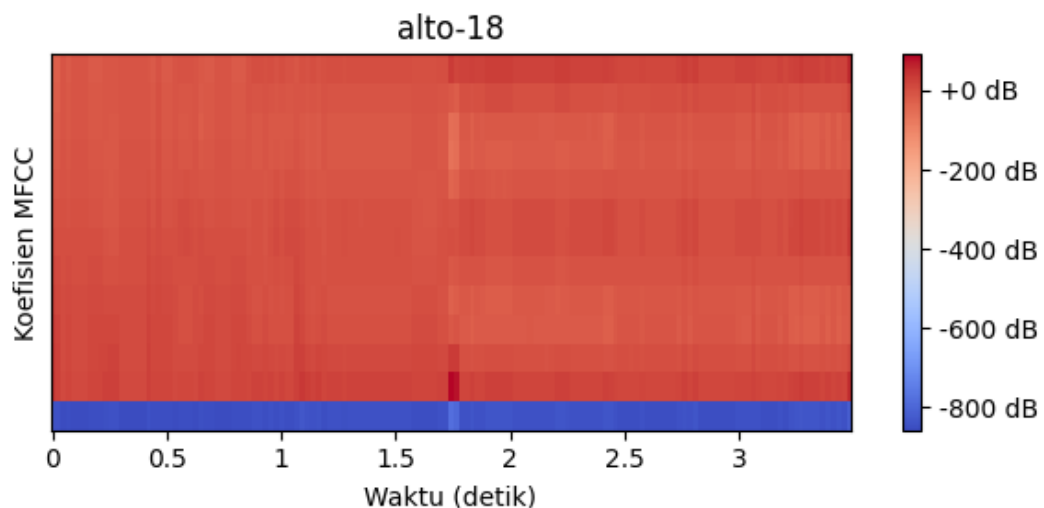
### IV.3 Evaluasi Subjektif Tahap Testing

Berdasarkan hasil pengujian, dilakukan evaluasi subjektif pada data yang mendapatkan prediksi yang salah dari model CRNN, CNN, ataupun RNN. Terdapat 3 data yang mengalami salah prediksi oleh CRNN, 6 data yang mengalami salah prediksi oleh CNN, dan 8 data yang mengalami salah prediksi oleh RNN. Totalnya terdapat 8 data yang perlu dilakukan evaluasi subjektif. Kelas Alto merupakan kelas yang paling banyak mengalami kesalahan pelabelan, sementara kelas Tenor merupakan kelas yang paling sedikit mengalami kesalahan pelabelan.



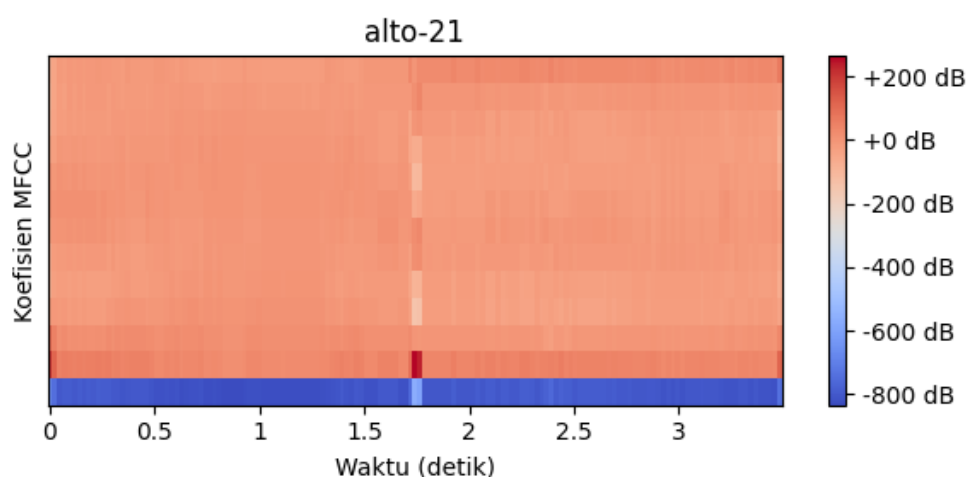
Gambar IV-19 MFCC alto-02

Pada audio alto-02, terlihat bahwa detik 1.75-3.5 memiliki lebih banyak warna gelap terutama di koefisien MFCC awal. Hal ini terjadi karena terdapat *noise* konstan pada detik tersebut. Selain itu, pada detik 0-1.75, terkadang terdapat *timestamp* yang memiliki warna yang kontras seperti pada detik 0.8, hal ini terjadi karena terdapat *noise* tidak konstan yang berasal dari *background* sehingga menyebabkan ketidakstabilan audio.



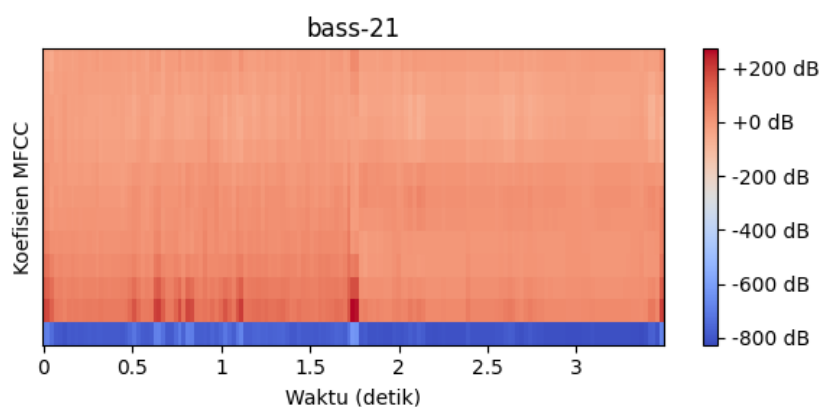
Gambar IV-20 MFCC alto-18

Pada audio alto-18, terdapat kemungkinan salah pelabelan audio. Hal ini terbukti dari *timbre* lebih stabil di nada C5, yang seharusnya merupakan "*sweet spot*" sopran.



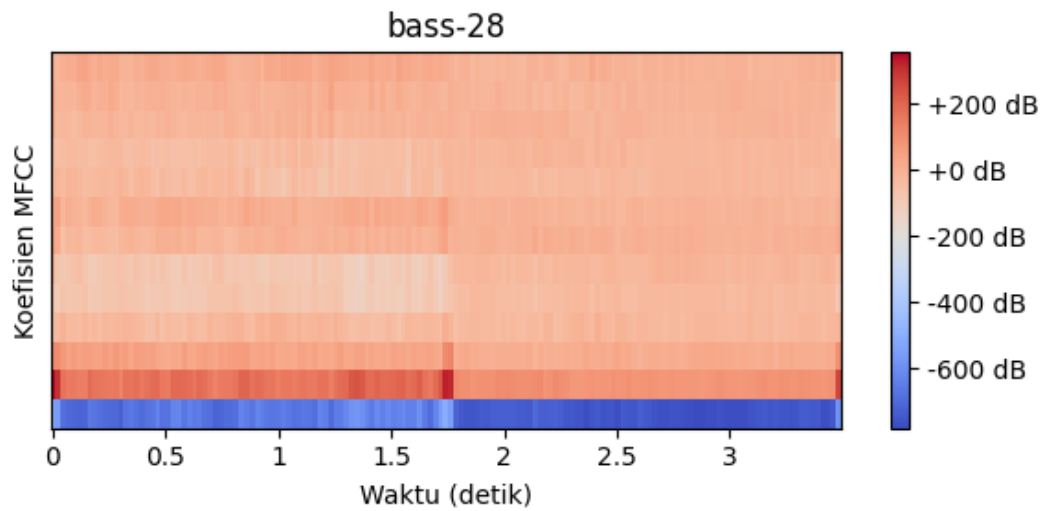
Gambar IV-21 MFCC alto-21

Pada audio alto-21, terdapat *noise* yang tidak konstan (suara listrik statis di awal) menyebabkan power yang ditangkap MFCC lebih rendah di awal, kemudian terdapat kemungkinan salah label, jika diperhatikan, *timbre* lebih stabil di nada C5, yang seharusnya merupakan "*sweet spot*" sopran.



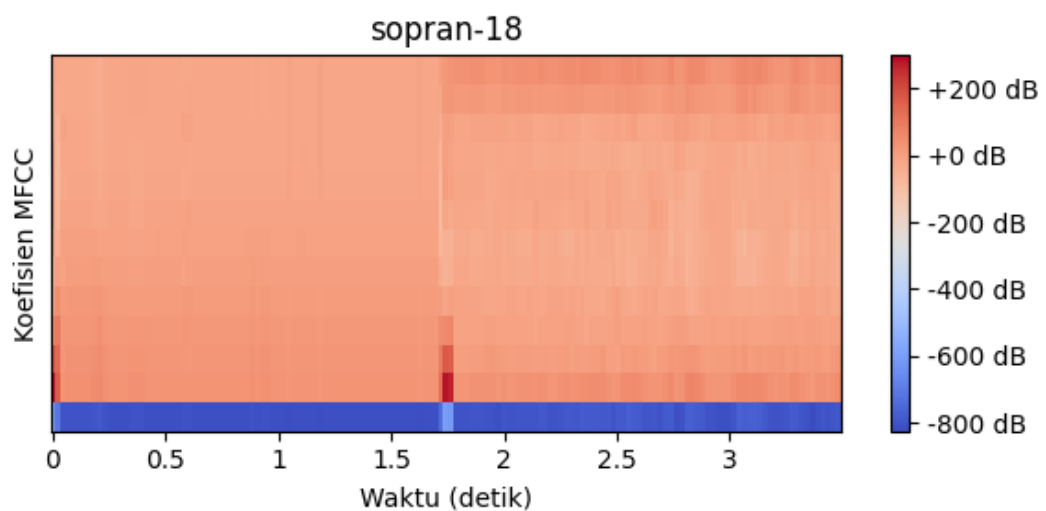
Gambar IV-22 MFCC bass-21

Pada audio bass-21, Clip mengalami rusak minor yang menyebabkan ketidakstabilan suara di awal.



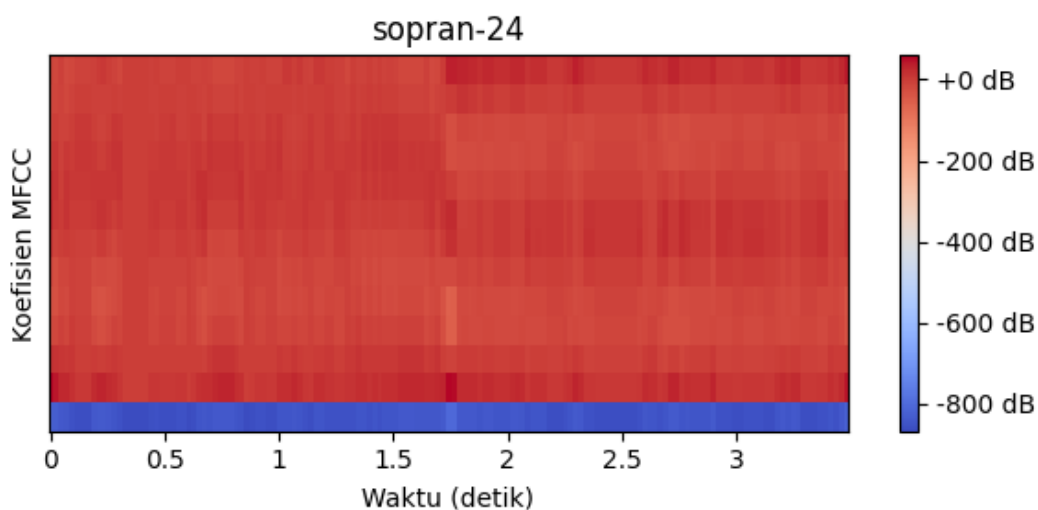
Gambar IV-23 MFCC bass-28

Pada audio bass-28, suara tidak terlalu stabil di awal. Terlihat dari *peak power* pada koefisien MFCC awal (1-3) relatif berantakan.



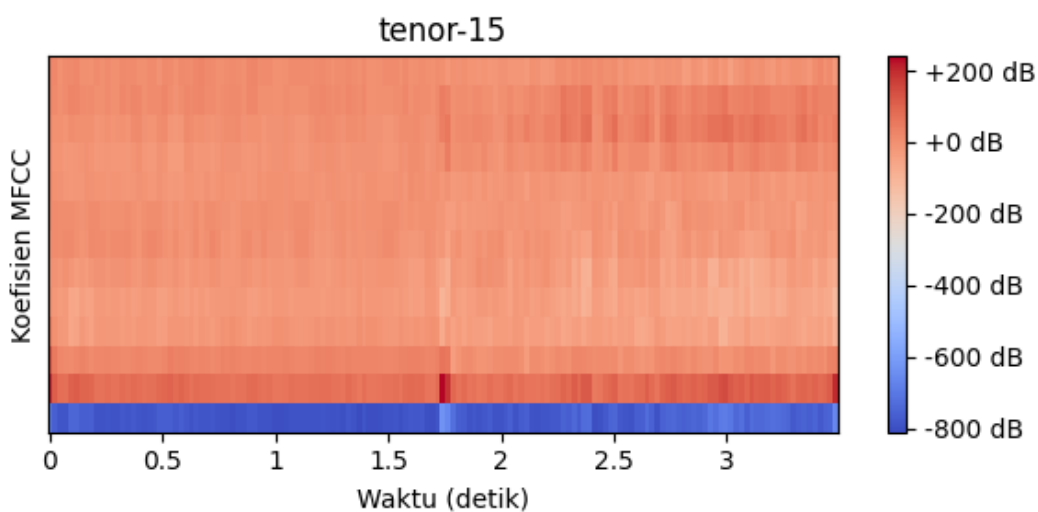
Gambar IV-24 MFCC sopran-18

Pada audio sopran-18, suara tidak stabil di akhir (*vibrato*), kemudian jika diperhatikan, *timbre* lebih stabil di nada C4, yang seharusnya merupakan "sweet spot" alto. Namun hal ini bisa saja terjadi karena pengaruh *vibrato*.



Gambar IV-25 MFCC sopran-24

Pada audio sopran-24, Clip mengalami rusak minor yang menyebabkan ketidakstabilan suara di akhir.



Gambar IV-26 MFCC tenor-15



Pada audio tenor-15, *Timbre* yang ada tidak terlalu wajar, yakni sangat stabil di akhir, kemungkinan karena terjadi *clipping* pada *device*. Selain itu juga terdapat *white noise* hanya di akhir clip, menyebabkan *power* yang tertangkap MFCC lebih tinggi daripada yang seharusnya.

#### **IV.4 Analisis Hasil Evaluasi Tahap Testing**

Berdasarkan hasil analisis model *convolutional recurrent neural network* serta model acuan (*convolutional neural network* dan *recurrent neural network*), dapat disimpulkan bahwa *convolutional recurrent neural network* memberikan hasil terbaik dari ketiga model. Hal ini menunjukkan bahwa hipotesis awal yakni menggabungkan *convolutional neural network* dan *recurrent neural network* terbukti menghasilkan model yang lebih baik dalam mengklasifikasikan data.

Namun perlu disadari bahwa sampel yang digunakan hanya 89 data latih dan 23 data uji, dan ini merupakan jumlah yang relatif kecil sehingga hasilnya masih jauh dari kata sempurna dalam mencerminkan kinerja model. Terdapat pula ketidakseragaman data, yakni data tidak direkam di studio dengan *device* yang sama sehingga aspek lingkungan tidak sepenuhnya dapat diabaikan.

## **BAB V**

### **KESIMPULAN DAN SARAN**

Bab Kesimpulan dan Saran bab terakhir sekaligus penutup dari laporan tugas akhir ini. Pada bab ini akan dibahas mengenai kesimpulan pengembangan model pembelajaran mesin pengklasifikasian golongan suara serta saran untuk penelitian kedepannya.

#### **V.1 Kesimpulan**

Berikut ini adalah beberapa hal yang dapat disimpulkan dari Tugas Akhir “Klasifikasi Golongan Suara dalam Paduan Suara dengan Menggunakan *Convolutional Recurrent Neural Network* (CRNN)”.

1. Berdasarkan hasil evaluasi yang sudah dilakukan, diketahui bahwa *Convolutional Recurrent Neural Network* berhasil memberikan hasil yang jauh lebih baik daripada model yang sudah ada sebelumnya yaitu *Convolutional Neural Network* dan *Recurrent Neural Network*. *Convolutional Recurrent Neural Network* relatif *robust* terhadap *noise* maupun ketidakstabilan suara.
2. Kemampuan *Convolutional Neural Network* memberikan hasil yang relatif baik, namun terdapat permasalahan ketika terdapat *noise* baik statis maupun dinamis.
3. Kemampuan *Recurrent Neural Network* relatif moderat dalam melakukan prediksi golongan suara, namun terdapat permasalahan ketika terdapat ketidakstabilan suara seperti *vibrato* maupun *noise* dinamis.

#### **V.2 Saran**

Berikut ini adalah beberapa saran dari Tugas Akhir “Klasifikasi Golongan Suara dalam Paduan Suara dengan Menggunakan *Convolutional Recurrent Neural*

*Network* (CRNN)” yang dapat mendukung penelitian kedepannya perihal pengklasifikasian golongan suara.

1. Meningkatkan kuantitas data. Pada penelitian ini hanya menggunakan 112 data secara keseluruhan. Jumlah ini relatif kecil jika dibandingkan dengan jumlah kelas yang ada.
2. Meningkatkan kualitas data. Pada penelitian ini, data diambil menggunakan perangkat yang berbeda-beda dan lingkungan yang berbeda-beda pula. Terlihat dari bagian IV-3, kinerja beberapa model sangat terpengaruh pada kualitas dari audio.
3. Melakukan *tuning* terhadap jenis *layer*. *Layer* yang digunakan dalam ketiga model dalam penelitian ini bersifat konstan untuk tiap eksperimennya. Melakukan *tuning* terhadap jenis *layer* mungkin dapat meningkatkan kinerja dari model.

## DAFTAR REFERENSI

- Anjali Pahwa, Gaurav Aggarwal. (2016). *Speech feature extraction for gender recognition*. International Journal of Image, Graphics and Signal Processing, 8(9), 17-25. doi:10.5815/ijigsp.2016.09.03
- Balabanovic, M. (1998). *Learning to surf: Multi-agent systems for adaptive web page recommendation*. Doctoral dissertation, Stanford University, Menlo Park, CA: Department of Computer Science.
- Elbir, A., Ilhan, H. O., Serbes, G., & Aydin, N. (2018). *Short Time Fourier Transform based music genre classification*. 2018 Electr. Electron. Comput. Sci. Biomed. Eng. Meet. EBBT 2018, 1-4. doi: 10.1109/EBBT.2018.8391437.
- Fisher, R. (2020). *An abridged choral director's guide to the male voice change*. *Music Educators Journal*, 107(1), 24–33. <https://doi.org/10.1177/8755123319890742>
- Hanna, J. R., & Deutsch, D. (2009). *The psychology of music (3rd ed.)*. San Diego, CA: Academic Press.
- Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques, third edition (3rd ed.)*. Morgan Kaufmann Publishers.
- Johns Hopkins Medicine. (2023). *Vocal cord disorders*. <https://www.hopkinsmedicine.org/health/conditions-and-diseases/vocal-cord-disorders>. Diakses pada 5 Juni 2023.
- Kennedy, M. (2007). *The Concise Oxford Dictionary of Music*. Oxford University Press. doi:10.1093/acref/9780199203833.001.0001
- McKusick, K.B., & Langley, P. (1991). *Constraints on tree structure in concept formation*. Prosiding The 21st ACM-SIGIR International Conference on Research and Development in Information Retrieval, 206-214. New York, NY:ACM Press.
- Nakano, T., Yoshii, K., Wu, Y., Nishikimi, R., Lin, K. W. Edward, & Goto, M. (2019). *Joint singing pitch estimation and voice separation based on a neural harmonic structure renderer*. 2019 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), 160-164. doi: 10.1109/WASPAA.2019.8937135.
- Octavya, Nanda Harsana, & Ubaya, Huda. (2021). *Implementasi model rekognisi suara menggunakan metode convolutional recurrent neural network (CRNN)*. Undergraduate thesis, Sriwijaya University.

- Pahwa, A., Aggarwal, G. (2016). *Speech feature extraction for gender recognition*. International Journal of Image, Graphics and Signal Processing, 8(9), 17-25. doi:10.5815/ijigsp.2016.09.03
- Pazzani, M., & Billsus, D. (1997). *Learning and revising user profiles: The identification of interesting web sites*. Machine Learning, 27, 313-331.
- Pavlichenko, Nikita, Stelmakh, Ivan, & Ustalov, Dmitry. (2021). CrowdSpeech and Vox DIY: Benchmark Dataset for Crowdsourced Audio Transcription [Data set]. Thirty-fifth Conference on Neural Information Processing Systems (NeurIPS 2021), Online. Zenodo. <https://doi.org/10.5281/zenodo.5574585>
- Pham, T. Van, Quang, N. T. N., & Thanh, T. M. (2019). *Deep learning approach for singer voice classification of Vietnamese popular music*. ACM Int. Conf. Proceeding Ser., 255-260. doi: 10.1145/3368926.3369700.
- Pratama, K. B., Suyanto, S., & Rachmawati, E. (2021). *Human vocal type classification using MFCC and convolutional neural network*. 2021 International Conference on Communication & Information Technology (ICICT), 43-48. doi: 10.1109/ICICT52195.2021.9568474.
- Ratmono, Wildo. (1985). *Pelajaran Seni Musik untuk SMA Kelas 1*. Surabaya: Sinar Wijaya
- Singh, N. (2020). *Classification of animal sound using convolutional neural network*. Masters Dissertation, Technological University Dublin. doi:10.21427/7pb8-9409

## **Lampiran A. Dataset**

*Dataset* yang digunakan dalam penelitian ini dapat diakses pada <https://github.com/stefanus-lamlo/Data-TA>

## Lampiran B. Hasil Prediksi Data Uji

Keterangan label:

0 = Sopran

1 = Alto

2 = Tenor

3 = Bass

Filename	True Class	CRNN	CNN	RNN
alto-02.wav	1	0	0	0
alto-03.wav	1	1	1	1
alto-08.wav	1	1	1	1
alto-15.wav	1	1	1	1
alto-18.wav	1	0	0	2
alto-21.wav	1	0	0	0
bass-05.wav	3	3	3	3
bass-14.wav	3	3	3	3
bass-19.wav	3	3	3	3
bass-21.wav	3	3	2	2
bass-28.wav	3	3	3	2
bass-29.wav	3	3	3	3
sopran-07.wav	0	0	0	0
sopran-08.wav	0	0	0	0
sopran-15.wav	0	0	0	0
sopran-18.wav	0	0	2	1

sopran-24.wav	0	0	0	2
sopran-26.wav	0	0	0	0
tenor-02.wav	2	2	2	2
tenor-15.wav	2	2	1	1
tenor-24.wav	2	2	2	2
tenor-26.wav	2	2	2	2
tenor-27.wav	2	2	2	2