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

Mei 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, 5 Mei 2023

Stefanus
NIM 13519101

ABSTRAK

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

Oleh

STEFANUS

NIM : 13519101

Abstrak berisi ringkasan apa yang telah dikerjakan dalam tugas akhir. Ada beberapa hal yang perlu diperhatikan dalam penulisan abstrak. Pertama, abstrak harus memuat permasalahan yang dikaji, metode/teknik yang digunakan untuk menyelesaikan masalah, hasil yang dicapai / evaluasi kajian, kesimpulan yang diperoleh, dan kata kunci. Kedua, cara penulisannya harus padat dan terarah. Setiap kalimat harus dapat memberikan informasi sebanyak dan setepat mungkin, mudah dibaca dan dimengerti. Panjang ringkasan dibatasi maksimal 300 kata dan ditulis dengan satu spasi. Panjang ringkasan dibatasi maksimal 300 kata dan ditulis dengan satu spasi.

Kata kunci: ringkasan, singkat, padat.

KATA PENGANTAR

Gunakan bagian ini untuk memberikan ucapan terima kasih kepada semua pihak yang secara langsung atau tidak langsung membantu penyelesaian tugas akhir, termasuk pemberi beasiswa jika ada. Utamakan untuk memberikan ucapan terima kasih kepada tim pembimbing tugas akhir dan staf pengajar atau pihak program studi, bahkan sebelum mengucapkan terima kasih kepada keluarga. Ucapan terima kasih sebaiknya bukan hanya menyebutkan nama orang saja, tetapi juga memberikan penjelasan bagaimana bentuk bantuan/dukungan yang diberikan. Gunakan bahasa yang baik dan sopan serta memberikan kesan yang enak untuk dibaca. Sebagai contoh: "Tidak lupa saya ucapkan terima kasih kepada teman dekat saya, Tito, yang sejak satu tahun terakhir ini selalu memberikan semangat dan mengingatkan saya apabila lengah dalam mengerjakan Tugas Akhir ini. Tito juga banyak membantu mengoreksi format dan *layout* tulisan. Apresiasi saya sampaikan kepada pemberi beasiswa, Yayasan Beasiswa, yang telah memberikan bantuan dana kuliah dan biaya hidup selama dua tahun. Bantuan dana tersebut sangat membantu saya untuk dapat lebih fokus dalam menyelesaikan pendidikan saya. ....". Ucapan permintaan maaf karena kekurangsempurnaan hasil Tugas Akhir tidak perlu ditulis.

DAFTAR ISI

DAFTAR LAMPIRAN

DAFTAR GAMBAR

DAFTAR TABEL

BAB I

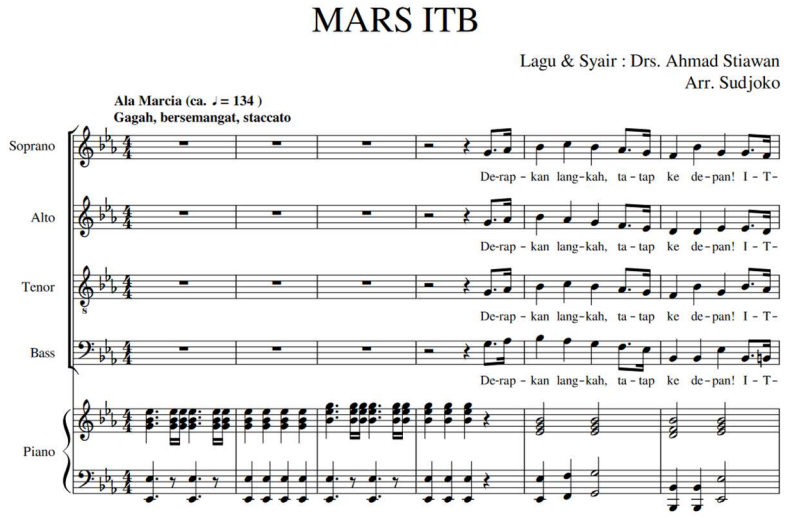
PENDAHULUAN

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).



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% (K. B. Pratama, 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. 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. Kondisi lingkungan untuk pemodelan data latih dan data uji seragam, sedemikian rupa sehingga tidak diperlukan adaptasi model akustik untuk menyesuaikan terhadap kondisi lingkungan.
2. Penyanyi yang dijadikan data latih dan data uji merupakan penyanyi yang dapat menyanyikan nada dengan tepat (tidak *tone deaf*).
3. *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 *cross-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

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 (K. B. Pratama, 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.

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 / f0*). Setiap manusia memiliki karakteristik suara masing-masing yang menyebabkan warna suara (*timbre*) yang berbeda-beda pula. Dalam *music information retrieval* (MIR), *timbre* sering kali disebut sebagai *harmonic series* (*f1,f2,f3,…,fn*) dengan rasio tiap frekuensi adalah N/N-1. Pada kenyataannya, teori *harmonic series* tidak selalu sesuai dengan kenyataan. Ada konsep yang biasa disebut sebagai *inharmonicity* dimana ada ketidaksesuaian antara frekuensi teori dengan frekuensi yang didapatkan di dunia nyata. Banyak aspek yang menyebabkan *inharmonicity*, misalnya faktor eksternal yang memengaruhi gelombang suara. Contohnya bila f0 = 440 hz (A4), maka f1 secara teori adalah 880 hz (A5). Namun karena *inharmonicity*, menyebabkan f1 menjadi 888 hz (Cano, E., dkk, 2018).

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	*Vocal Range* (Nada)	*Vocal Range* (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

II.3 Metode Klasifikasi Golongan Suara

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 lampau, didapatkan bahwasanya mayoritas metode yang digunakan adalah menggunakan *Artificial Neural Network* (ANN) dimana metode ini cukup umum digunakan. (Pahwa, A, dkk, 2016). Pada jurnal lainnya, juga digunakan *Mel-Frequency Cepstral Coefficient* (MFCC) sebagai *feature extractor* dan menganalisis hasilnya menggunakan *Support Vector Machine* (SVM) serta *Convolutional Neural Network* (CNN) sebagai *classifier*-nya. Penggunaan *Mel-Frequency Cepstral Coefficient* terbukti meningkatkan akurasi dari pengklasifikasian jenis kelamin (Pahwa, A, dkk, 2016). Sejalan dengan pengklasifikasian suara penyanyi dalam musik populer di Vietnam oleh Toan Pham Van pada 2019, Penggunaan MFCC dan CNN menghasilkan presisi yang cukup baik (93%) (Van, T.P., dkk, 2019).

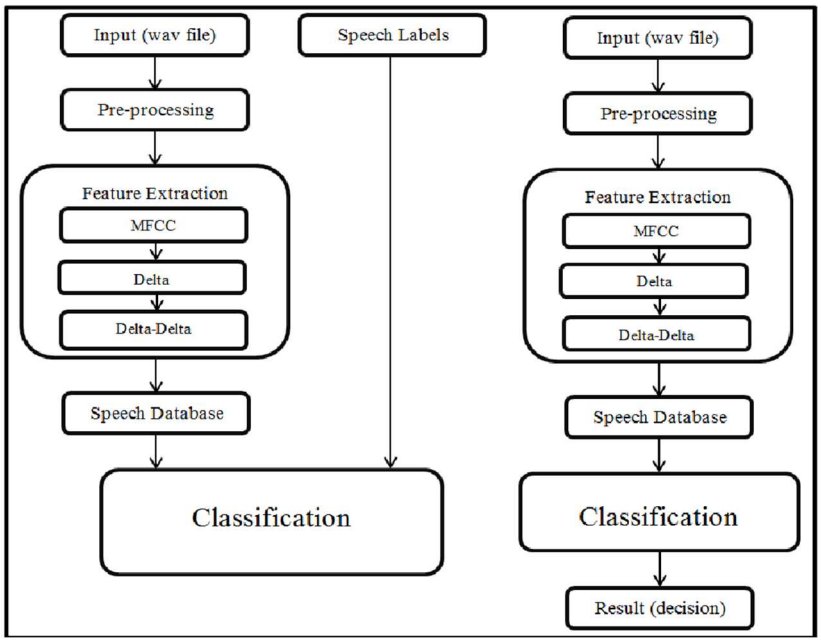
Dari studi-studi tersebut, terbukti bahwasanya *Convolutional Neural Network* (CNN) mampu menghasilkan hasil yang cukup baik dalam klasifikasi data audio. Fokus utama dalam pemrosesan data audio adalah pengekstrakan fitur yang

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) 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* yang sudah cukup baik dalam pemrosesan data audio. Secara teoritis, *convolutional recurrent neural network* memiliki keunggulan dalam pemrosesan sinyal audio.

II.3.1 Metode Klasifikasi Suara Menggunakan *Artificial Neural Network*

Untuk melakukan klasifikasi menggunakan *Artificial Neural Network* (ANN), berdasarkan gambar II-1, terdapat beberapa tahap besar yang harus dilakukan, yakni *feature extraction* dan *classification*. Tahapan *feature extraction* digunakan untuk mentransformasi data sinyal audio ke dalam bentuk representasi yang lebih stabil. Ada banyak jenis fitur yang dapat diekstrak dalam *feature extraction* seperti *pitch*, *energy*, *Mel-Frequency Cepstral Coefficient* (MFCC), *delta,* dan *delta-delta*. Tahapan *classification* meliputi kegiatan *training*, *testing*, dan *data validation*. Penggunaan *stacking* antara *Support Vector Machine* dan *Artificial Neural Network* digunakan dalam tahapan ini. (Pahwa, A, dkk, 2016)



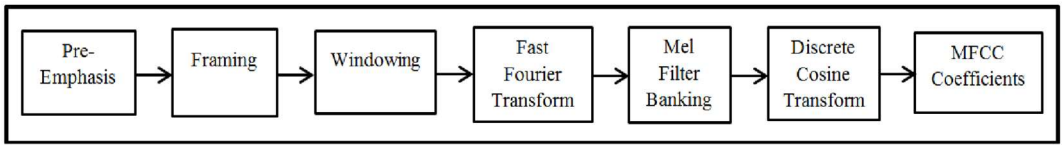
Gambar II-1 Arsitektur Pengklasifikasian Suara Menggunakan *Artificial Neural Network*
(Pahwa, A, dkk, 2016)

II.3.1.1 *Feature Extraction*

Pada dasarnya, *feature extraction* adalah proses pengubahan sinyal audio menjadi bentuk representasi numerikal. Dalam kasus *gender classification*, ada banyak jenis fitur yang dapat diekstrak, namun berdasarkan pengamatan, fitur yang paling memengaruhi adalah *Mel-Frequency Cepstral Coefficient* (Pahwa, A, dkk, 2016). *Mel-Frequency Cepstral Coefficient* (MFCC) meniru cara kerja telinga manusia. MFCC memiliki beberapa tahap seperti *pre-emphasis*, *framing*, *windowing*, *fast fourier transform*, *mel filter bank*, *discrete cosine transform*, hingga menghasilkan *MFCC coefficients*.

Seperti yang digambarkan dalam gambar II-2, dalam tahap *pre-emphasis*, sinyal yang masuk melalui suatu filter yang akan meningkatkan frekuensi tinggi dalam sinyal audio. Pada tahap *framing*, sinyal audio dipisah-pisahkan ke dalam blok-blok tersendiri yang disebut juga dengan kata *frame*. Sementara pada tahap *windowing*, tiap *frame* dikalikan dengan suatu nilai *hamming window*. Setelah itu, *frame* masuk ke tahap *fast fourier transform* dimana dalam tahapan ini, dilakukan transformasi

dari domain waktu ke domain frekuensi karena dalam bentuk domain frekuensi, data audio lebih mudah dianalisis. Tahapan selanjutnya adalah nilai domain frekuensi dilakukan *mapping* terhadap *mel frequency*. *Mel frequency* adalah frekuensi subjektif yang meniru cara kerja pendengaran manusia. Tahap terakhir adalah *discrete cosine transform* yang mana mengubah bentuk *mel frequency* ke dalam bentuk *mel-frequency cepstral coefficient* (MFCC).



Gambar II-2 Langkah *Mel-Frequency Cepstral Coefficient*
(Pahwa, A, dkk, 2016)

II.3.1.2 *Classification*

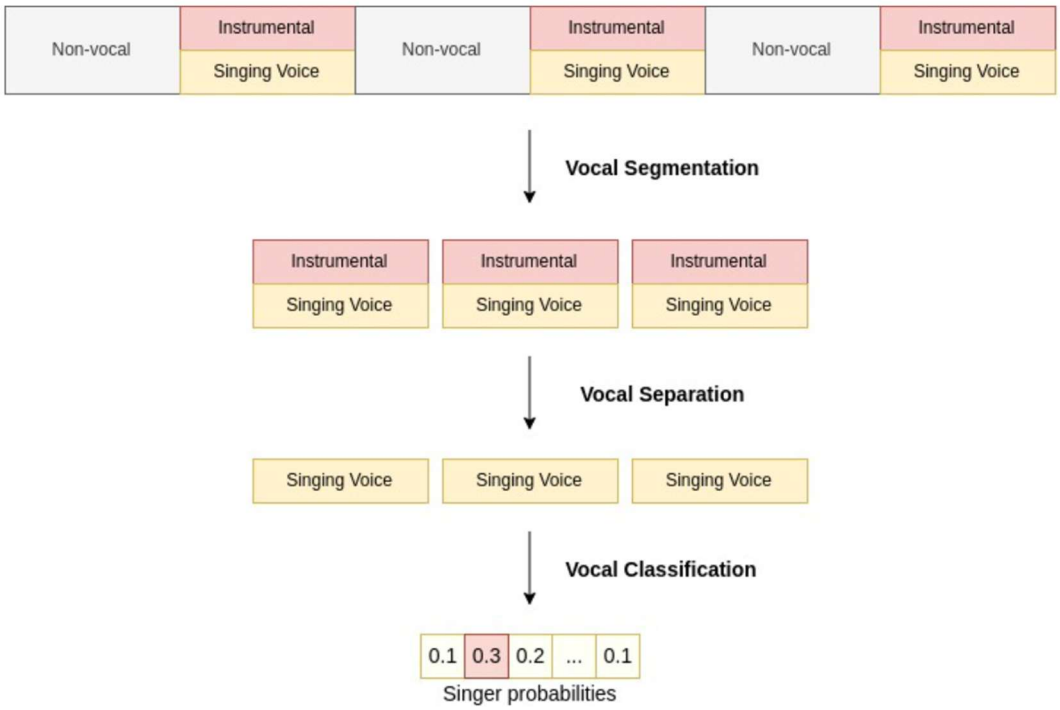
Tahapan klasifikasi meliputi tahapan *training* dan *testing* dari model menggunakan sebuah kakas bantu bernama RapidMiner. RapidMiner adalah kakas yang dibuat oleh perusahaan RapidMiner yang memiliki berbagai fungsi dalam lingkup *data science* seperti *data mining* serta *business analysis*. Kakas RapidMiner menerima berbagai macam jenis basis data sebagai masukan. Terdapat banyak pilihan *classifier* di dalam RapidMiner untuk dipilih, pada kasus ini yang dipilih adalah *Artificial Neural Network* (ANN). Alasan pemilihan ANN sebagai *classifier* adalah karena ANN adalah algoritma dengan riset yang cukup mendalam. Selain itu ANN juga merupakan algoritma yang dapat berkembang sendiri berdasarkan data masukan tanpa fungsi maupun model khusus.

II.3.2 Metode Klasifikasi Suara Menggunakan *Convolutional Neural Network*

Berdasarkan gambar II-3, dalam pengklasifikasian suara menggunakan *Convolutional Neural Network*, Terdapat beberapa tahapan yang perlu dilakukan,

yakni *vocal segmentation*, *vocal separation*, dan *vocal classification*. Untuk tiap tahapan, digunakan arsitektur *CNN* yang berbeda dengan *dataset* yang berbeda pula dengan tujuan memaksimalkan akurasi dari sistem yang dibuat. (Van, T.P., dkk, 2019)

Vocal segmentation memiliki tujuan utama yaitu mengidentifikasikan suara vokal dan suara lingkungan. Data yang dihasilkan oleh *vocal segmentation* kemudian digunakan untuk melakukan *vocal separation* untuk mengekstrak vokal dari suara lingkungannya. Fase ini sangat penting bagi akurasi dari hasil prediksi. Setelah suara vokal sudah terisolasi, data ini digunakan untuk melakukan klasifikasi.



Gambar II-3 Proses pengklasifikasian suara menggunakan *Convolutional Neural Network*
(Van, T.P., dkk, 2019)

II.3.2.1 *Vocal Segmentation*

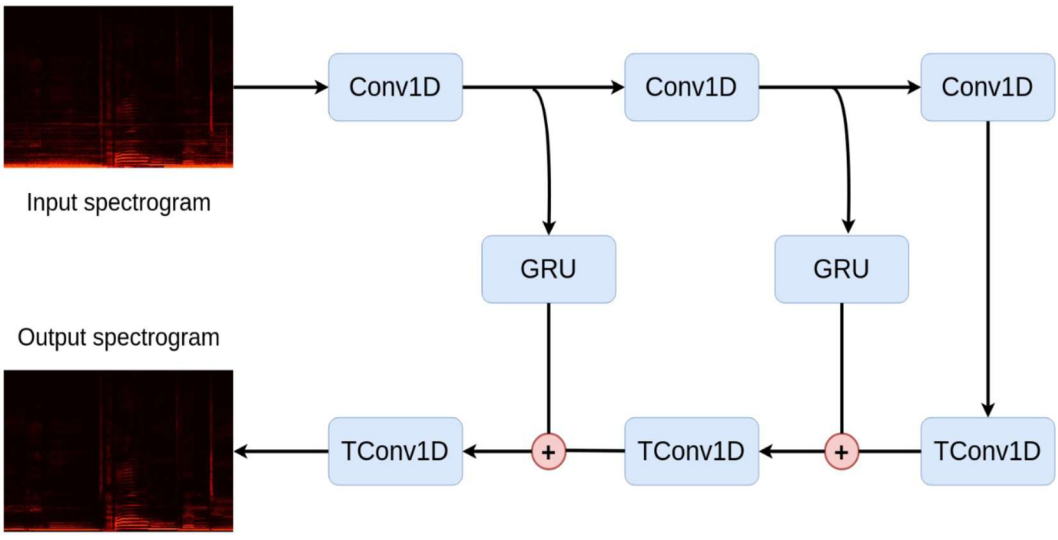
Tujuan utama dalam *vocal segmentation* adalah untuk mendeteksi batasan dalam sinyal audio untuk memisahkan suara vokal dengan suara non-vokal. Pemanfaatan

convolutional neural network untuk mendeteksi batasan sinyal audio ini menggunakan *dataset* berupa musik karaoke berdasarkan 50 *Mel-Frequency Cepstral Coefficient* yang merepresentasikan 500 milisekon data audio. Output yang dikeluarkan berupa 2 neuron yang merepresentasikan vokal dan non-vokal.

II.3.2.2 *Vocal Separation*

Berdasarkan gambar II-4, tahapan *vocal separation* dimulai dengan membagi audio menjadi potongan sepanjang 6 sekon yang kemudian diproses melalui *Short-Time Fourier Transform* (STFT). Setelah melalui STFT, akan dihasilkan properti frekuensi temporal untuk tiap *timeframe* yang didefinisikan panjangnya. Hasil *magnitude matrix* ini akan dinormalisasi menggunakan skala logaritma. Secara intuitif, hal ini akan menghasilkan suatu spektogram dalam suatu skala desibel.

Spektogram yang dihasilkan ini kemudian dimasukkan ke dalam *Convolutional Neural Network* melalui beberapa proses termasuk menggunakan *Gated Recurrent Unit* (GRU). Keluaran dari model ini adalah spektogram vokal yang telah terpisahkan dari latar belakangnya.



Gambar II-4 Arsitektur Model *Vocal Separation* dengan pemanfaatan *Convolutional Neural Network*
(Van, T.P., dkk, 2019)

II.3.2.3 *Vocal Classification*

Berdasarkan spektogram vokal yang dihasilkan, dilakukan ekstraksi *Mel-Frequency Cepstral Coefficients* (MFCC). MFCC ini kemudiannya akan dimasukkan ke dalam model pengklasifikasian suara. Pada dasarnya MFCC menekankan kepada frekuensi rendah yang merepresentasikan manusia daripada frekuensi-frekuensi lainnya yang tidak terlalu informatif. Hasil MFCC ini kemudiannya dimasukkan ke dalam model LSTM, lebih tepatnya *bidirectional LSTM* dengan 3 *hidden layer* yang berguna untuk melakukan klasifikasi vokal.

II.4 *Convolutional Recurrent Neural Network*

II.5 Hasil Penelitian Terkait

Berdasarkan penelitian Pahwa, A, 2016, dengan menggunakan *Artificial Neural Network* dengan model yang dijelaskan pada bagian II.3.1 dalam penggolongan jenis kelamin berdasarkan sinyal suara, didapatkan hasil akurasi 96% untuk jenis kelamin perempuan dan 90.48% untuk jenis kelamin laki-laki. Hal ini menandakan penggunaan *Artificial Neural Network* dengan akurasi tertentu dapat digunakan dalam pengklasifikasian golongan suara biner (laki-laki atau perempuan).

Sementara berdasarkan penelitian Van Pham, Toan, 2019, dengan menggunakan *Convolutional Neural Network* dengan model yang juga telah dijelaskan pada bagian II.3.2, dalam penggolongan penyanyi di Vietnam, didapatkan bahwa penggunaan *Convolutional Neural Network* dengan analisis fitur *Mel-Frequency Cepstral Coefficient*, didapatkan hasil berupa nilai presisi 93%. Hal ini menandakan bahwa penggunaan *Convolutional Neural Network* dengan nilai presisi tertentu dapat digunakan dalam pengklasifikasian golongan suara *multi-label* dengan *target class* penyanyi itu sendiri.

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, seperti yang dijelaskan pada bagian II.4, akurasi pada penggunaan *Convolutional Recurrent Neural Network* menghasilkan *training accuracy* sebesar 99.41% serta *testing accuracy* sebesar 99.05%. Hal ini menandakan penggunaan *Convolutional Recurrent Neural Network* menghasilkan akurasi yang lebih baik dibandingkan model-model sebelumnya.

II.6 Metrik Evaluasi

Metrik evaluasi yang digunakan dalam klasifikasi golongan suara adalah metrik evaluasi standar yang sering digunakan dalam klasifikasi. Berdasarkan Jiawei Han (2012), Dalam klasifikasi, metrik evaluasi yang sering digunakan adalah akurasi (*accuracy*), yaitu persentase kebenaran prediksi model dibandingkan dengan jumlah total data yang digunakan untuk evaluasi.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+F}} \tag{2.1}$$

Selain itu, metrik evaluasi lain yang sering digunakan dalam klasifikasi adalah presisi (*precision*) dan sesitivitas (*recall*). *Precision* mengukur seberapa sering model memprediksi suatu kelas dengan benar, sedangkan *recall* mengukur seberapa baik model dapat menemukan semua contoh dari suatu kelas.

$$Precision = \frac{TP}{TP+FP} \tag{2.2}$$

$$Recall = \frac{TP}{TP+FN} \tag{2.3}$$

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*.

$$F1 = \frac{2*Precision*Recall}{Precision+Rec} = \frac{2*TP}{2*TP+FP+FN} \tag{2.4}$$

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

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* (*f0*) karena salah satu karakteristik dari *Convolutional Neural Network* adalah kemampuannya dalam mendeteksi hubungan antar fitur

yang berdekatan. Hal ini sejalan dengan penelitian yang sebelumnya dilakukan oleh K. B. Pratama pada 2021. Sementara untuk kestabilan suara bisa dideteksi dengan data masukan yang memiliki *timeseries* contohnya dengan mengunakan 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*. Hal ini sejalan dengan penelitian yang dilakukan Toan Pham Van di tahun 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.

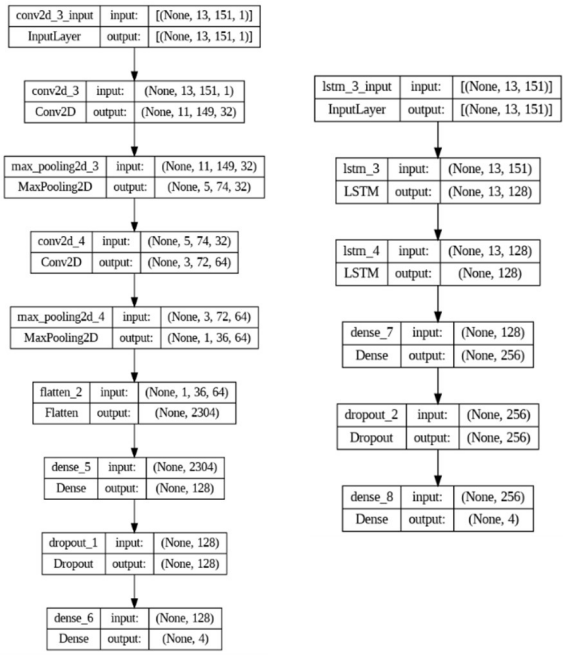
Salah satu model arsitektur yang dapat dipertimbangkan untuk mengatasi permasalahan yang ada adalah menggunakan arsitektur *Convolutional Recurrent Neural Network* (CRNN). CRNN merupakan arsitektur gabungan antara *convolutional neural network* dan *recurrent neural network* yang mana menggabungkan kelebihan 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.

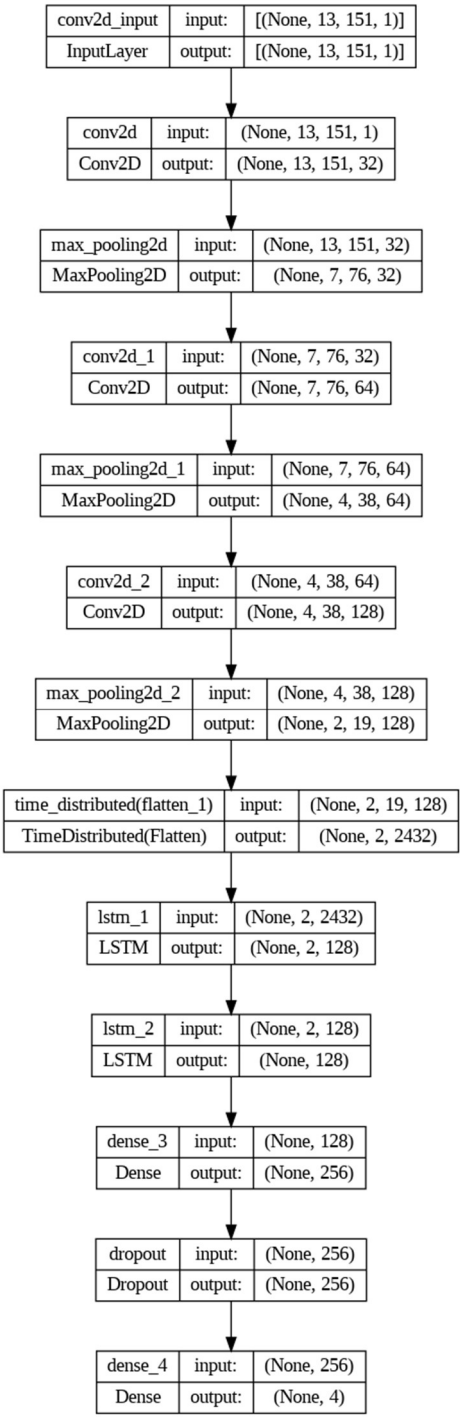
III.3 Rancangan Solusi

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 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 esktraksi 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 berfungsikan 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 pertujuan 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*.

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*. *Tuning* yang dilakukan adalah menggunakan *epoch=[100,200,500]* dan *learning_rate=[0.1,0.001,0.0005,0.0001,0.00005,0.00001]*. Berdasarkan hasil eksperimen, diambil *history* dari masing-masing eksperimen dan disimpan ke dalam suatu tabel, berikut adalah tabel yang dimaksud.

Tabel IV-1 Hasil Eksperimen

Experiment	Accuracy	Val_Accuracy
epoch_100_lr_0.01	0.571428597	0
epoch_100_lr_0.001	1	0.1538461596
epoch_100_lr_0.0001	0.9795918465	0.4615384638
epoch_100_lr_0.0005	1	0.1538461596

epoch_100_lr_0.00001	0.5510203838	0.6923077106
epoch_100_lr_0.00005	0.9387755394	0
epoch_200_lr_0.01	0.5918367505	0
epoch_200_lr_0.001	1	0.3076923192
epoch_200_lr_0.0001	1	0
epoch_200_lr_0.0005	1	0.1538461596
epoch_200_lr_0.00001	0.7755101919	0.1538461596
epoch_200_lr_0.00005	0.9795918465	0
epoch_500_lr_0.01	0.571428597	0
epoch_500_lr_0.001	1	0.2307692319
epoch_500_lr_0.0001	1	0.07692307979
epoch_500_lr_0.0005	1	0.1538461596
epoch_500_lr_0.00001	0.9795918465	0.07692307979
epoch_500_lr_0.00005	1	0

Berdasarkan tabel IV-1, terlihat sering sekali terjadi *overfitting* dimana *accuracy* 1 namun *val_accuracy* 0. Hal ini berarti *accuracy* sangat baik bagi *data training* namun buruk untuk *data testing*. Sehingga berdasarkan rata-rata dari *accuracy dan val_accuracy*, *hyperparameter* terbaik didapatkan pada *epoch* 100 dan *learning rate* 0.0001.

IV.2 Evaluasi Tahap Testing

Berdasarkan evaluasi tahap *training*, dibuatlah suatu model baru menggunakan *hyperparameter* terbaik yang sudah didapatkan, yakni *epoch 100* dan *learning_rate*

0.0001. Model dilatih dengan seluruh 62 data *training*. Kemudian dibandingkan dengan model acuan (CNN dan CRNN). Berikut merupakan *classification report* bawaan *sklearn* untuk CRNN (tabel IV-2), CNN (tabel IV-3) dan RNN (tabel IV-4).

Tabel IV-2 *Classification Report CRNN*

	Precision	Recall	F1-score	Support
Alto	0.93	1	0.97	14
Bass	0.67	0.62	0.65	13
Sopran	0.82	0.88	0.85	16
Tenor	0.75	0.67	0.71	18
accuracy			0.76	61
macro avg	0.79	0.79	0.79	61
weighted avg	0.77	0.76	0.76	61

Tabel IV-3 *Classification Report CNN*

	Precision	Recall	F1-score	Support
Alto	0.29	1	0.45	14
Bass	0	0	0	13
Sopran	0	0	0	16
Tenor	1	0.06	0.11	18
accuracy			0.29	61
macro avg	0.32	0.26	0.14	61
weighted avg	0.34	0.29	0.13	61

Tabel IV-4 *Classification Report RNN*

	Precision	Recall	F1-score	Support
Alto	0.5	0.29	0.36	14
Bass	0.31	0.31	0.31	13
Sopran	0.5	0.25	0.33	16
Tenor	0.47	0.78	0.58	18
accuracy			0.44	61
macro avg	0.44	0.41	0.4	61
weighted avg	0.45	0.44	0.42	61

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 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 61 data, 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*.
2. Kemampuan *Convolutional Neural Network* memberikan hasil yang relatif buruk, dimana hampir seluruh dari prediksinya mengembalikan kelas Alto. Hal ini disebabkan ketidakmampuan *convolutional neural network* untuk mengambil fitur *timeseries* dimana keterurutan data tidak dapat ditinjau.
3. Kemampuan *Recurrent Neural Network* relatif moderat dalam melakukan prediksi golongan suara, hal ini disebabkan warna suara seseorang sebenarnya dapat berubah sewaktu-waktu, sedangkan hal yang dianggap paling penting adalah *range* suara. Hal ini lebih tertinjau melalui fitur *timeseries*.

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. Sebaiknya TA dikerjakan!

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.

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