

BAB II TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

2.1.1 A Siren Identification System Using Deep Learning To Aid Hearing-Impaired People

Penelitian yang disajikan dalam jurnal ini bertujuan untuk mengatasi *safety issue* yang dihadapi oleh orang-orang dengan gangguan pendengaran dalam mengidentifikasi suara sirene dengan sebuah sistem identifikasi sirene menggunakan sistem *deep learning* yang dirancang, dibangun, dan diuji. Sistem ini terdiri dari *neural network* tiruan dengan menggunakan teknik pengenalan gambar untuk mengidentifikasi keberadaan sirene dengan mengubah suara yang masuk ke dalam spektogram. Permasalahan pada penelitian ini yaitu kurangnya dataset untuk pelatihan jaringan yang dapat diatasi dengan cara menghasilkan data yang sesuai dari berbagai suara sirene yang dicampur dengan kebisingan lingkungan secara relevan. *Hardware interface* juga dikembangkan untuk mengkomunikasikan deteksi sirene dengan pengguna, menggunakan metode visual. Setelah pelatihan model, sistem diuji secara menyeluruh menggunakan skenario realistis untuk mengevaluasi kinerjanya. Untuk suara sirene yang digunakan selama pelatihan, sistem berhasil mencapai tingkat akurasi sebesar 98 persen. Sementara itu, untuk suara sirene di dunia nyata yang direkam di jalan-jalan pusat kota London, akurasinya mencapai 91 persen. Saat diuji di lingkungan yang bising, hasil menunjukkan bahwa sistem mampu mendeteksi suara sirene meskipun berada pada tingkat -6 dB di bawah kebisingan latar. Hasil ini menunjukkan bahwa sistem yang dikembangkan dapat menjadi dasar bagi aplikasi deteksi sirene untuk membantu penyandang gangguan pendengaran.

Berdasarkan penelitian tersebut, terdapat poin – poin yang menjadi acuan dalam penelitian yang dilakukan oleh peneliti yaitu sebagai berikut:[6]

1) Penggunaan CNN dan Spektrogram untuk Deteksi Suara Sirene

Convolutional Neural Network (CNN) terbukti efektif dalam mendeteksi sirene ambulans dengan cara mengonversi sinyal audio menjadi spektrogram. Metode ini memungkinkan CNN untuk mengekstrak fitur dari

audio sebagai gambar, sehingga meningkatkan akurasi deteksi sirene dalam lingkungan dengan tingkat kebisingan tinggi.

2) Implementasi pada *Embedded Device* untuk Responsivitas Real-Time

Penelitian tersebut juga menggarisbawahi pentingnya implementasi sistem berbasis CNN pada *embedded device* seperti *microcontroller* atau prosesor berdaya rendah. Hal ini memungkinkan deteksi sirene ambulans dilakukan secara *real-time* dengan konsumsi daya rendah. Implementasi ini cocok untuk perangkat yang dipasang pada kendaraan atau digunakan langsung oleh pengemudi, sehingga respons terhadap suara sirene dapat dilakukan dengan cepat

3) Penerapan Sinyal Peringatan Visual untuk Pengguna dengan Gangguan Pendengaran

Penelitian Ramirez et al. menyarankan penggunaan indikator visual atau getaran untuk memberikan peringatan ketika sirene terdeteksi, khususnya bagi pengemudi dengan gangguan pendengaran. Sistem yang dikembangkan memanfaatkan antarmuka fisik dengan lampu LED untuk memberikan sinyal visual kepada pengguna bahwa sirene darurat telah terdeteksi, yang sangat relevan jika diterapkan pada penelitian ini.

4) Validasi Model dalam Lingkungan Nyata

Penelitian tersebut melakukan uji coba pada kondisi perkotaan di London untuk mengevaluasi performa model dalam lalu lintas sesungguhnya. Simulasi atau uji lapangan serupa bisa dilakukan pada penelitian ini untuk memastikan sistem mampu mendeteksi suara sirene ambulans secara akurat dalam lingkungan yang kompleks dan dinamis.

2.1.2 Implementasi CNN Untuk Deteksi Suara Pada Sirene Ambulance

Pendeteksian suara sirene ambulans memiliki peran penting dalam berbagai aplikasi, termasuk pengembangan sistem peringatan dini dan bantuan navigasi kendaraan darurat. Penelitian ini mengkaji implementasi *Convolutional Neural Network* (CNN) untuk mendeteksi suara sirene ambulans secara otomatis. CNN dipilih karena kemampuannya yang unggul dalam mengenali pola dari data suara yang kompleks. Dataset yang digunakan dalam penelitian ini mencakup berbagai

rekaman suara sirine ambulans yang diambil dalam kondisi lingkungan yang berbeda-beda. Data ini kemudian diproses menjadi spektrogram untuk digunakan sebagai input ke dalam jaringan CNN. Hasil penelitian menunjukkan bahwa model CNN yang diusulkan mampu mendeteksi suara sirine ambulans dengan akurasi yang tinggi, mencapai rata-rata akurasi sebesar 95%. Implementasi ini juga memperlihatkan ketahanan yang baik terhadap variasi noise lingkungan, menjadikannya solusi yang andal untuk aplikasi real-time. Dengan demikian, penelitian ini membuktikan bahwa CNN dapat menjadi alat yang efektif untuk mendeteksi suara sirine ambulans, yang berpotensi meningkatkan responsivitas sistem peringatan dan navigasi darurat.

Berdasarkan penelitian tersebut, terdapat poin – poin yang menjadi acuan dalam penelitian yang dilakukan oleh peneliti yaitu sebagai berikut:[11]

1) Penggunaan *Convolutional Neural Network* (CNN)

Dalam penelitian tersebut menunjukkan bahwa CNN dapat mendeteksi suara sirene ambulans dengan tingkat akurasi lebih dari 90%, bahkan dalam kondisi noise yang bervariasi. Selain itu, CNN dikenal efektif dalam mengenali pola suara yang kompleks dan menggunakan metode Mel Spectrogram untuk filtrasi suara.

2) Perbandingan Model

Penelitian tersebut membandingkan efisiensi CNN dengan model tradisional seperti BRANN dan model lain yang menggunakan YOLOv8 pada perangkat keras berdaya tinggi.

3) Portabilitas dan Skalabilitas

Model CNN yang lebih ringan dan dapat diterapkan pada perangkat *embedded* seperti ponsel, hal tersebut dapat menawarkan keuntungan dalam portabilitas dan skalabilitas

2.1.3 Acoustic Based Emergency Vehicle Detection Using Ensemble of deep Learning Models

Dalam penelitian ini, *emergency vehicle detection* (EVD) seperti pemadam kebakaran, ambulans, dan Struktur temporal dan spektral dimiliki dalam domain

waktu-frekuensi oleh peristiwa suara. Menganalisis dan mengklasifikasikan lingkungan akustik menggunakan perekaman suara adalah ruang lingkup penelitian yang sedang berkembang. Lapisan konvolusi dapat dengan cepat mengekstrak fitur tingkat tinggi dan menggeser fitur varian dari *time-frequency domain*. Dalam penelitian ini, dataset diambil dari Google Audioset ontology yang selanjutnya dikumpulkan dan fitur-fiturnya diekstraksi dengan *Mel-frequency Cepstral Coefficient* (MFCC). Tiga sistem model yang diteliti yaitu *deep neural networks* (DNN), *Convolutional Neural Network* (CNN), dan *Recurrent Neural Network* (RNN) dengan konfigurasi dan parameter yang berbeda telah diteliti. Kemudian, sebuah ensemble model telah dirancang dengan parameter yang dipilih secara optimal.

Berdasarkan penelitian tersebut, terdapat poin – poin yang menjadi acuan dalam penelitian yang dilakukan oleh peneliti yaitu sebagai berikut:[12]

1) Ekstraksi Fitur Menggunakan MFCC

Mel Frequency Cepstral Coefficients (MFCC) digunakan sebagai metode ekstraksi fitur utama untuk suara sirene. MFCC dikenal sebagai pendekatan yang sangat baik untuk menangkap karakteristik frekuensi dari suara sirene, karena mampu memberikan representasi kompak dari spektrum sinyal audio. Dalam penelitian ini, kombinasi CNN dan MFCC menghasilkan hasil yang optimal dalam klasifikasi suara sirene, yang bisa diadaptasi dalam penelitian Anda untuk meningkatkan sensitivitas deteksi.

2) Pengembangan Dataset dengan Kebisingan Variatif

Dataset yang beragam dapat meningkatkan ketahanan dan akurasi model dalam mengenali suara sirene ambulans, sehingga sistem mampu mendeteksi suara darurat meskipun terdapat suara latar yang kuat.

2.1.4 Ambulance Siren Detection using Artificial Intelligence in Urban Scenarios

Dalam penelitian ini, dataset sirene ambulans dan suara lalu lintas telah dikembangkan dalam format '.wav'. Sinyal audio yang dikumpulkan diproses terlebih dahulu sebelum fitur diekstraksi menggunakan *Python Librosa library*. Dua puluh satu *Mel-frequency Cepstral Coefficient* (MFCC), *spectral centroid*, *roll*

of rate, spectral bandwidth, Chroma_stft, dan zero-crossing rate diekstraksi. GPU digunakan untuk melatih model MLP, yang mencapai tingkat akurasi sebesar 90% selama proses pelatihan. Fokus dari penelitian ini adalah pada perancangan dataset dan proses ekstraksi fitur, dengan tujuan membedakan suara ambulans dari kebisingan jalan yang paling signifikan. Namun, peneliti belum melakukan pengujian data dalam lingkungan dunia nyata. Peneliti merencanakan pengujian data secara *real-time* di masa depan dengan merekam lebih banyak dataset audio. Model yang dikembangkan dapat diterapkan dalam praktik setelah dataset ditingkatkan dan diuji dalam skenario dunia nyata. Model yang telah dilatih ini dapat diprogram pada perangkat *microcontroller* apa pun dan diimplementasikan dalam aplikasi nyata untuk mendeteksi sirene ambulans serta mengoptimalkan waktu respons. Dari segi kompleksitas, model MLP yang ada cukup sederhana untuk dijalankan pada *micro-processing gadget* atau panel apa pun tanpa memerlukan *graphical processing unit*.

Berdasarkan penelitian tersebut, terdapat poin – poin yang menjadi acuan dalam penelitian yang dilakukan oleh peneliti yaitu sebagai berikut:[13]

1) Pemanfaatan CNN dan Spektrogram untuk Pendeteksian Sirene

Convolutional Neural Network (CNN) mampu melakukan deteksi suara sirene ambulans dengan mengonversi sinyal audio menjadi bentuk spektrogram. Teknik ini memungkinkan CNN untuk menangkap fitur suara melalui representasi visual, sehingga meningkatkan ketepatan dalam mengenali suara sirene bahkan di lingkungan dengan kebisingan tinggi. Hal ini memungkinkan sistem mendeteksi sirene ambulans secara akurat meskipun ada gangguan suara latar yang intens dari lalu lintas.

2) Pengembangan Dataset yang Melibatkan Berbagai Tingkat Kebisingan

Dalam penelitian tersebut menunjukkan bahwa memiliki dataset yang mengandung variasi kebisingan dan jenis sirene dari lingkungan nyata, seperti suasana lalu lintas kota, dapat meningkatkan ketangguhan model. Dataset yang luas dengan karakteristik suara yang beragam memungkinkan sistem untuk mendeteksi suara darurat dengan lebih baik meskipun dalam kondisi suara latar yang ramai.

2.2 Tinjauan Teori

2.2.1 *Machine Learning*

Machine learning merujuk pada sebuah metode yang membuat komputer memiliki kemampuan dalam mempelajari dan melakukan sebuah pekerjaan secara otomatis. Proses *machine learning* dilakukan melalui algoritma tertentu, sehingga pekerjaan yang diperintahkan kepada komputer dapat dilakukan secara otomatis [14]. Menurut Mohri et al., [15] *machine learning* dapat didefinisikan sebagai metode komputasi berdasarkan pengalaman untuk meningkatkan performa atau membuat prediksi yang akurat. Definisi pengalaman disini ialah informasi sebelumnya yang telah tersedia dan bisa dijadikan data pembelajaran.

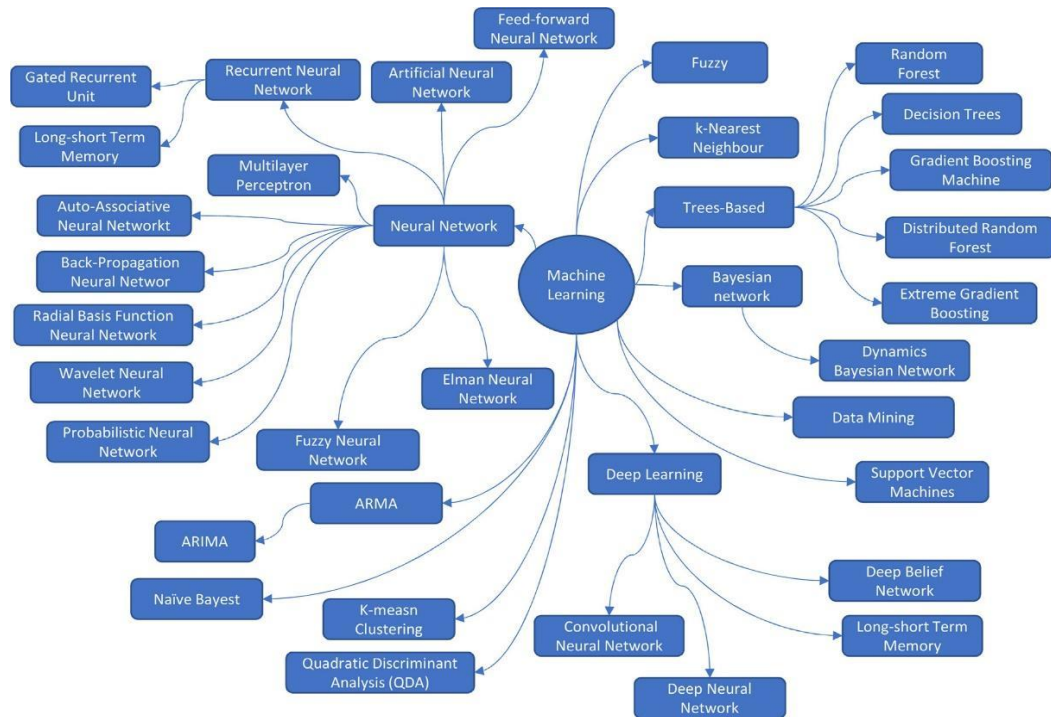
Metode *machine learning* membuat sistem desain yang dapat mempelajari perilaku dalam sebuah dataset. Metode ini memungkinkan sistem untuk belajar dan berkembang berdasarkan pengalaman dan waktu yang merujuk pada model yang dapat memprediksi hasil dari pembelajaran data lampau [16]. Mesin dapat mempelajari perilaku berdasarkan tiga faktor [17], yaitu:

- 1) Data masa lampau yang telah dipelajari oleh mesin
- 2) Matriks yang dapat mengukur *error* dari perilaku aktual dan ideal
- 3) Mekanisme timbal balik yang mengukur *error* untuk membuat program menghasilkan perilaku yang lebih baik kedepannya

Algoritma dalam *machine learning* tersebut terbagi menjadi 3 (tiga) tipe yaitu sebagai berikut:[17]

- 1) *Supervised learning* adalah pembelajaran mesin berdasarkan data historis yang memiliki setiap set input dan output. Salah satu contoh yang termasuk kedalam *supervised learning* adalah klasifikasi.
- 2) *Unsupervised learning* adalah pembelajaran mesin berdasarkan data historis yang setiap input tidak dilengkapi dengan output. Salah satu contoh yang termasuk kedalam *unsupervised learning* adalah klastering.
- 3) *Reinforcement learning* adalah tipe pembelajaran yang diperlakukan terpisah dari *supervised* dan *unsupervised*. Tipe ini memerlukan timbal balik dari lingkungan sehingga tidak sepenuhnya *unsupervised*. Namun, data yang ada tidak selalu memiliki data yang berlabel selama pelatihan mesin

sehingga tidak bisa dimasukkan ke dalam kategori *supervised*. Sistem secara terus menerus akan berinteraksi dengan lingkungan dalam pencairan untuk menghasilkan perilaku yang diinginkan dan mendapat timbal balik dari lingkungan.



Gambar 2. 1 Gambar *Taxonomy of Machine Learning Techniques*

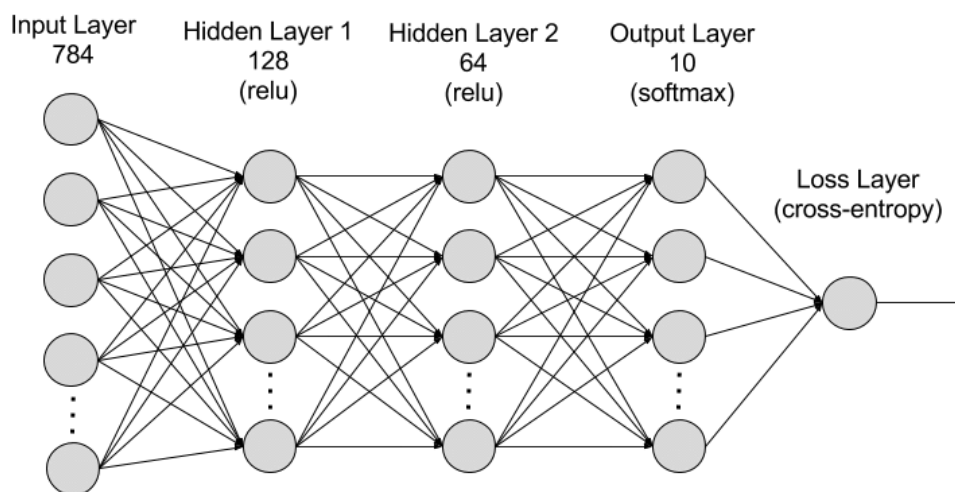
Berdasarkan gambar 1 tersebut, dapat dilihat bahwa *Convolutional Neural Network* (CNN) bisa digunakan untuk memproses spectrogram. *Convolutional Neural Network* CNN akan mempelajari fitur-fitur penting dari spectrogram untuk mendeteksi pola suara sirene.

2.2.2 *Deep Learning*

Deep Learning merupakan sub-bidang dari *Machine Learning* yang bekerja dengan melakukan representasi dari data pada lapisan pembelajaran yang berlapis-lapis hingga representasi menjadi lebih bermakna. *Deep* pada *Deep Learning* memiliki makna bahwa *Deep Learning* menggagas lapisan representasi yang berurutan. Sekarang ini, *Deep Learning* biasanya memiliki lapisan dengan jumlah

puluhan hingga ratusan lapisan yang berurutan dan lapisan-lapisan tersebut dapat secara otomatis mempelajari data *training* yang diberikan. Pada *Deep Learning* lapisan representasi tersebut bernama *Neural Networks*. *Neural Networks* memiliki struktur yang bertumpuk yang berarti suatu lapisan akan berada diatas lapisan lain. *Neural Networks* mengambil konsep dari bidang *neurobiology* dan terinspirasi dari kemampuan dalam memahami sesuatu seperti yang dilakukan oleh otak manusia. Meskipun konsep utama *Deep Learning* dikembangkan atas inspirasi tersebut, tetapi model *Deep Learning* bukanlah model dari otak manusia [18].

Teknik *Deep Learning* memberikan arsitektur yang sangat kuat untuk *Supervised learning*. Model pembelajaran dapat dengan menambah lapisan lebih banyak dapat mewakili data citra berlabel yang lebih baik. *Deep Learning* banyak digunakan pada komunitas penelitian dan industry dalam membantu pemecahan masalah dengan data yang besar seperti *Computer Vision*, *Speech Recognition*, dan *Natural Language Processing* [19].



Gambar 2. 2 Gambar Cara Kerja *Deep Learning*

Algoritme *deep learning* merupakan jaringan neural yang meniru otak manusia. Misalnya, otak manusia memiliki jutaan neuron yang saling terhubung yang bekerja sama untuk mempelajari dan memproses informasi. Demikian pula, jaringan *neural deep learning*, atau jaringan neural buatan, terbuat dari banyak lapisan neuron buatan yang bekerja sama di dalam komputer.

Dalam *deep learning*, metode CNN atau *Convolutional Neural Network* sangatlah bagus dalam menemukan fitur yang baik pada citra ke lapisan berikutnya

untuk membentuk hipotesis nonlinear yang dapat meningkatkan kekompleksitasan sebuah model. Model yang kompleks tentunya akan membutuhkan waktu pelatihan yang lama sehingga di dunia *deep learning* penggunaan GPU sudah sangatlah umum [20].

2.2.3 Convolutional Neural Network (CNN)

Convolutional neural network (CNN) adalah salah satu algoritma *deep learning* yang dirancang untuk mengolah data dua dimensi. CNN biasanya digunakan untuk mempelajari dan mendeteksi *feature* pada sebuah gambar 2. Jika *neural network* pada umumnya menggunakan data *array* 1 dimensi sebagai *input*, pada algoritma CNN data yang digunakan sebagai *input* adalah data dua dimensi. Sama seperti *neural network* pada umumnya, CNN juga terdiri dari banyak *neuron* yang memiliki *weight*, *bias*, dan *activation function*. Convolutional neural network memiliki enam lapisan pembangun, yaitu: *convolutional layer*, *pooling layer*, *normalization layer*, *Rectified Linear Units layer*, *fully connected layer*, dan *loss layer* [21].

Convolutional layer merupakan lapisan inti pada CNN. Pada lapis ini seluruh bidang reseptif akan ditelusuri oleh filter secara tumpang tindih parsial. Penelusuran filter secara tumpang tindih parsial menyebabkan setiap neuron akan berbagi bobot koneksi (*weight sharing*). *Pooling layer* merupakan lapisan untuk meringkas output dari setiap grup neuron pada lokasi kernel yang sama. *Pooling layer* dilakukan untuk menjaga ukuran data ketika melakukan konvolusi dengan cara melakukan *down sampling*. Dengan lapis ini, representasi data menjadi lebih kecil, mudah dikelola, dan mudah dalam mengendalikan *overfitting*. *Normalization layer* digunakan untuk melakukan normalisasi pada citra masukan. Normalisasi pada citra masukan bertujuan untuk mengatasi perbedaan rentang nilai yang signifikan. *Rectified Linear Units (ReLU)* layer berfungsi untuk meningkatkan sifat nonlinearitas fungsi keputusan dan jaringan secara keseluruhan tanpa mempengaruhi bidang reseptif pada lapis konvolusi. *Fully connected layer* merupakan lapis yang memiliki kesamaan dengan *Multi Layer Perceptron (MLP)*. Pada lapis ini dilakukan perkalian matriks yang diikuti dengan bias offset [21].

Pada penelitian ini akan dilakukan Klasifikasi suara sirene ambulans menggunakan metode *Convolutional Neural Network* (CNN) dengan menggunakan ekstraksi fitur *Mel-Frequency Cepstrum Coefficient* (MFCC). *Mel-Frequency Cepstrum Coefficient* atau biasa disebut dengan MFCC adalah metode ekstraksi fitur yang populer di bidang teknologi suara [22]. Dalam proses ini, rekaman suara instrumen musik tiup dikonversi menjadi matriks konvolusi, yang dikenal sebagai spektrogram, atau dalam bentuk sinyal suara. Selanjutnya, fitur-fitur ini dapat digunakan untuk melakukan klasifikasi suara instrumen musik tiup yang terkandung di dalamnya.

Setelah melakukan ekstraksi fitur akan dilakukan proses klasifikasi menggunakan metode *Convolutional Neural Network* atau biasa dikenal dengan CNN adalah metode *deep learning* yang terutama menggunakan filter seperti matriks ukuran kecil untuk mengekstrak fitur dari kumpulan data yang besar. ConvNet dari CNN dianggap lebih mampu dalam menangkap ketergantungan spasial dan temporal jika dibandingkan dengan model-model lainnya [23]. CNN dapat digunakan untuk berbagai tugas seperti pengenalan wajah, pengolahan teks menjadi digital, dan pemrosesan bahasa alami, yang diyakini memiliki keunggulan yang signifikan

Penggunaan metode CNN memiliki beberapa keunggulan utama yang cocok untuk melakukan klasifikasi suara instrumen musik tiup. CNN mampu mengekstrak fitur dari data yang sangat kompleks melalui penggunaan lapisan konvolusi yang memungkinkan model untuk menangkap pola yang lebih halus dalam data suara yang mungkin terlewatkan oleh metode lain. Dengan MFCC sebagai input, CNN dapat lebih efektif mengidentifikasi karakteristik unik dari setiap instrumen musik tiup. Suara memiliki struktur temporal yang kompleks dan CNN sangat efektif dalam menangani data yang memiliki hubungan spasial dan temporal [24].

2.2.4 *Mel-Frequency Cepstral Coefficients* (MFCC)

Dalam pengenalan suara, terdapat dua proses yaitu proses ekstraksi dan identifikasi. Hal ini juga terjadi dalam identifikasi nada yang dihasilkan oleh suara sirene. Beberapa metode yang sering digunakan dalam ekstraksi sinyal suara adalah *Mel-Frequency Cepstrum Coefficients* (MFCC) [25], Short-Term Fourier

Transform (STFT), Linear Predictive Coding (LPC), dan Wavelet Transform. MFCC dipilih karena kemampuannya untuk memberikan representasi kompak dari spektrum audio yang sesuai dengan sensitivitas pendengaran manusia. Hal ini membuat MFCC sangat efektif dalam menangkap karakteristik frekuensi dari suara sirine ambulans dibanding metode lain. Tahapan ekstraksi fitur MFCC terdiri dari *Pre-emphasis*, *Windowing*, *Fast Fourier Transform*, *Mel Filter Bank*, *Discrete Cosine Transform*, dan *Delta Feature* [26] yang dijelaskan sebagai berikut:

1) *Pre-emphasis*

Pada tahap ini sinyal suara akan disaring. Metode penyaringan ini mengurangi nilai frekuensi sinyal, hanya sinyal frekuensi tinggi yang dapat lolos penyaringan, pada langkah ini noise akan dikurangi, sehingga hanya data sinyal suara yang sebenarnya yang dapat ditangkap oleh sistem.

2) *Windowing*

Fungsi *windowing* untuk meminimalkan diskontinuitas pada frame awal dan akhir. Sebagian besar penelitian menggunakan hamming window karena kesederhanaan rumus dan nilai kerja *windowing*. Windowing digunakan untuk meminimalkan diskontinuitas pada frame awal dan akhir sinyal audio dengan mengurangi efek sinyal diskret yang tidak kontinu. Jendela seperti Hamming Window mengurangi amplitudo sinyal di ujung frame, sehingga menghasilkan transisi yang lebih halus dan mencegah distorsi frekuensi yang tidak diinginkan pada proses Fourier Transform.

$$w(n) = 0.54 - 0.45 \cos\left(\frac{2\pi n}{N-1}\right) \quad 0 \leq n \leq N-1$$

Equation 1. *Windowing*

Ekuasi 1. diatas menggambarkan fungsi jendela Hanning yang digunakan untuk mengurangi efek kebocoran dalam analisis sinyal, dengan memberikan bobot lebih pada bagian tengah sinyal dan mengurangi bobot pada ujungnya, di mana n adalah indeks diskrit antara 0 hingga $N-1$, dan N adalah panjang sinyal.

3) *Fast Fourier Transform*

FFT adalah algoritma yang mengimplementasikan *Discrete Fourier Transform* (DFT), DFT mengubah setiap frame dari domain waktu ke domain frekuensi.

$$X_k = \sum_{n=0}^{N-1} X_n e^{-j2\frac{\pi}{N}kn}$$

Equation 2. *Fast Fourier Transform*

Ekuasi 2. diatas menggambarkan transformasi Fourier diskrit (DFT) yang menghitung komponen frekuensi X_k dari sinyal diskrit X_n dengan mengalikan nilai sinyal X_n dengan faktor eksponensial kompleks untuk setiap indeks frekuensi K.

4) *Mel Filter Bank*

Frekuensi linier yang diperoleh dari FFT diubah menjadi skala *Mel-Frequency* untuk mendapatkan limit bank filter berdasarkan Skala Frekuensi Mel.

$$F_{mei} = 2595 \log\left(1 + \frac{f}{700}\right)$$

Equation 3. *Mel Filter Bank*

Ekuasi 3. diatas menggambarkan konversi frekuensi f (dalam Hz) menjadi satuan mel (unit pengukuran skala persepsi frekuensi), di mana F_{mei} adalah nilai mel dan f adalah frekuensi dalam Hertz, yang digunakan untuk menggambarkan persepsi pendengaran manusia terhadap perubahan frekuensi.

5) *Discrete Cosine Transform*

Hasil log dari domain waktu DCT disebut MFCC. Persamaan MFCC dapat dilihat pada persamaan berikut:

$$X_k = \sum_{n=0}^{N-1} 2x[n] \cos\left(\frac{n}{2N}k + (2n + 1)\right)$$

Equation 4. *Discrete Cosine Transform*

Ekuasi 4. diatas menggambarkan bentuk transformasi Fourier atau analisis sinyal yang menggunakan komponen kosinus untuk menghitung nilai X_k dengan $x[n]$ sebagai data input, di mana rumus ini melibatkan faktor-faktor yang berhubungan dengan frekuensi diskrit k dan indeks n , serta modifikasi pada fase dan amplitudo sinyal.

6) *Delta Feature*

Hasil perhitungan delta akan ditambahkan ke vektor ciri, sehingga menghasilkan vektor ciri yang lebih besar untuk meningkatkan akurasi sistem *Automated Speech Recognition* (ASR). Metode ini akan menghasilkan koefisien delta sebesar koefisien cepstral yang dihasilkan oleh MFCC [27].

$$d_t = \frac{\sum_{n=1}^N n(C_{t+n} - C_{t-n})}{2 \sum_{n=1}^N n^2}$$

Equation 5. *Delta Feature*

Ekuasi 5. diatas menggambarkan perhitungan selisih terarah (derivative) dari fungsi C_t ada titik waktu t , di mana C_{t+n} dan C_{t-n} adalah nilai fungsi C_t pada $t + n$ dan $t - n$, dan n adalah faktor pengali yang diakumulasikan pada jumlah terbobot untuk menghitung perubahan relatif terhadap waktu, dengan pembobotan kuadrat n^2 pada penyebut untuk normalisasi.