

DETEKSI NADA DASAR GITAR MENGGUNAKAN FAST FOURIER TRANSFORM (FFT) DAN CONVOLUTIONAL NEURAL NETWORK (CNN)

ANALICIA, ATIK WINTARTI, YUNI ROSITA DEWI*

Program Studi Sarjana Sains Data, Universitas Negeri Surabaya, Indonesia

Abstrak

Penelitian ini dilakukan dengan tujuan mendeteksi nada dasar gitar menggunakan pendekatan kuantitatif. Pengumpulan data dilakukan dengan melakukan pengolahan sinyal dari setiap nada dasar menggunakan dataset Kaggle yang memiliki 43 nada dasar, setiap kelas memiliki minimal 24 file audio dari satu nada dasar gitar. Penelitian ini menggunakan metode *Fast Fourier Transform* (FFT) sebagai ekstraksi fitur dan *Convolutional Neural Network* (CNN) sebagai klasifikasi model terhadap tingkat akurasi sehingga mampu mengevaluasi hasil deteksi nada dasar gitar. Praproses penelitian ini dilakukan dengan menghilangkan noise, melakukan *high-pass filter*, *pre-emphasis*, normalisasi amplitudo, dan ekstraksi fitur FFT untuk memperoleh representasi frekuensi tiap nada. Fitur spectral ini digunakan sebagai input bagi model 1D-CNN yang dirancang untuk mengklasifikasi nada dasar. Satuan akurasi untuk model deteksi nada dasar gitar menggunakan FFT dan CNN, yaitu evaluation matrix, model menghasilkan akurasi 99.53%, Precision 99.66%, Recall 99.60%, dan F1-Score 99.60%. Hasil uji coba terhadap beberapa lagu mencapai akurasi tertinggi 95%. Dengan demikian, penelitian ini efektif untuk deteksi nada dasar gitar.

Kata Kunci *Deteksi Nada Dasar; Sinyal Digital; Fast fourier Transform (FFT); Convolutional Neural Network (CNN)*

Abstract

This study was conducted to detect the fundamental tone of guitar using a quantitative approach. Data collection was carried out by processing signals from each fundamental tone using Kaggle dataset. containing 43 fundamental tones, each class contains at least 24 audio files of a single guitar fundamental tone. This study used the Fast Fourier Transform (FFT) method for feature extraction and Convolutional Neural Network (CNN) for model classification to determine the level of accuracy so that it can evaluate the results of guitar fundamental tone detection. The preprocessing involved noise removal, high-pass filtering, pre-emphasis, amplitude normalization, and FFT feature extraction to obtain the frequency representation of each tone. These spectral features were used as input for the 1D-CNN model designed to classify fundamental tones. Modelling using FFT and CNN resulted in an Accuracy of 99.53, Precision 99.66, Recall 99.60, F1-Score 99.60. The results on several songs achieved an accuracy rate of up to 95%. This research demonstrates that the proposed method is effective for detecting guitar fundamental tone.

Keywords *Fundamental Frequency Detection; Digital Signal; Fast Fourier Transform (FFT);*

*Corresponding author. Email: yunidewi@unesa.ac.id.

Convolutional Neural Network (CNN)

1 Pendahuluan

Gitar merupakan alat musik yang memiliki peran penting dalam dunia musik dan seni, serta menjadi salah satu instrumen yang paling populer di dunia, digunakan baik dalam pertunjukan profesional maupun sebagai media belajar musik bagi pemula. Ketepatan nada yang dihasilkan gitar sangat bergantung pada frekuensi suara tiap senarnya, di mana setiap senar memiliki nada dasar standar yang khas agar menghasilkan bunyi yang harmonis ketika dimainkan bersama senar lain maupun instrumen lain dalam sebuah ansambel. Penyetelan nada dasar secara manual menggunakan telinga sering tidak konsisten karena keterbatasan pendengaran manusia dalam membedakan selisih frekuensi yang kecil, sehingga hasilnya dapat bervariasi antar individu tergantung pengalaman dan ketajaman pendengaran masing-masing. Kesalahan kecil dalam penyetelan dapat membuat suara senar terdengar *false* dan tidak nyaman didengar, bahkan dapat mengganggu keseluruhan harmoni saat gitar dimainkan bersama instrumen lain. Permasalahan ini menjadi lebih signifikan bagi pemain gitar pemula yang belum terlatih mendengar perbedaan nada secara presisi, maupun dalam konteks produksi audio yang menuntut akurasi tinggi. Oleh karena itu, dibutuhkan metode komputasi yang mampu mendeteksi nada dasar gitar secara cepat, konsisten, dan akurat tanpa bergantung sepenuhnya pada kemampuan pendengaran manusia (Stefani and Turchet, 2025).

Secara umum, terdapat dua pendekatan utama dalam mendeteksi nada suatu sinyal audio: pendekatan berbasis pengolahan sinyal klasik dan pendekatan berbasis pembelajaran mesin. Pendekatan pertama biasanya mengandalkan analisis domain frekuensi secara langsung, sedangkan pendekatan kedua memanfaatkan model yang dilatih untuk mengenali pola dari data. Sinyal suara gitar sendiri bersifat periodik dan mengandung informasi frekuensi fundamental beserta harmoniknya, sehingga representasi frekuensi menjadi kunci penting dalam proses deteksi nada. *Fast Fourier Transform* (FFT) merupakan algoritma efisien untuk mengubah sinyal dari domain waktu ke domain frekuensi, sehingga memudahkan identifikasi pola frekuensi suatu nada (Cooley and Tukey, 1965). Namun, FFT memiliki keterbatasan: ia tidak memiliki informasi waktu kemunculan suatu frekuensi, cukup sensitif terhadap *noise* pada sinyal, dan pada praktiknya keputusan klasifikasi nada masih memerlukan aturan ambang batas (*threshold*) manual yang kurang adaptif terhadap variasi karakteristik sinyal antar instrumen atau kondisi rekaman. Di sisi lain, perkembangan kecerdasan buatan, khususnya *Convolutional Neural Network* (CNN), memberikan pendekatan baru dalam mengenali pola spektrum frekuensi dari representasi sinyal audio, termasuk hasil transformasi Fourier satu dimensi (1D) (Kow and Kow, 2025). CNN mampu mempelajari pola-pola diskriminatif secara otomatis dari data tanpa memerlukan aturan ambang batas manual, sehingga berpotensi lebih adaptif terhadap variasi sinyal dibandingkan pendekatan berbasis aturan. Kombinasi FFT dan CNN dinilai saling melengkapi: FFT mengubah sinyal ke domain frekuensi sehingga pola harmonik nada menjadi lebih mudah dikenali, sedangkan CNN mempelajari dan mengenali pola harmonik dari representasi frekuensi tersebut secara otomatis dan adaptif.

Beberapa penelitian terdahulu telah memanfaatkan FFT maupun CNN pada domain audio, namun dengan fokus yang berbeda-beda. Wijaya and Adikara (2025) menggunakan FFT untuk mendeteksi suara senar instrumen musik berbasis web, namun akurasinya masih dipengaruhi oleh kondisi akustik ruangan dan *noise*, sehingga performanya cenderung menurun pada lingkungan perekaman yang tidak terkontrol. Jannereth and Esch (2021) memanfaatkan FFT un-

tuk menganalisis timbre berbagai instrumen musik, dengan fokus pada karakterisasi warna suara alih-alih deteksi nada secara spesifik. Sementara itu, Kow and Kow (2025) menggunakan FFT sebagai representasi input bagi *Convolutional Autoencoder* untuk meningkatkan kualitas sinyal audio, menunjukkan bahwa kombinasi representasi frekuensi dengan arsitektur berbasis konvolusi dapat memberikan hasil yang baik pada tugas pemrosesan sinyal audio secara umum. Pada sisi klasifikasi berbasis CNN, Lei (2022) menggunakan CNN untuk mengklasifikasikan efek gitar (seperti distorsi atau chorus) dari sinyal audio, dan Pratama et al. (2023) menerapkan 1D-CNN untuk estimasi pitch suara penyanyi, keduanya menunjukkan bahwa CNN mampu menangkap karakteristik spektral yang relevan untuk tugas klasifikasi audio meskipun pada domain yang berbeda dari deteksi nada dasar gitar. Studi-studi tersebut menunjukkan bahwa FFT efektif sebagai metode ekstraksi fitur frekuensi, sementara CNN efektif dalam mengenali pola spektral, tetapi keduanya jarang digunakan secara bersamaan yang secara spesifik ditujukan untuk deteksi nada dasar gitar akustik, dan sebagian besar penelitian yang ada juga hanya diuji pada kondisi data terkontrol tanpa mengevaluasi performanya pada nada-nada yang muncul dalam konteks lagu yang dimainkan secara nyata.

Berdasarkan uraian tersebut, terdapat gap penelitian yang jelas: belum banyak penelitian yang menggabungkan FFT sebagai ekstraksi fitur dan CNN sebagai model klasifikasi secara khusus untuk mendeteksi nada dasar gitar akustik, khususnya yang diuji tidak hanya pada data nada tunggal terkontrol, tetapi juga pada nada-nada yang muncul dalam lagu nyata dengan kondisi akustik yang lebih bervariasi. Penelitian ini berupaya mengisi kesenjangan tersebut dengan mengombinasikan kedua metode tersebut secara *end-to-end*, mulai dari praproses sinyal, ekstraksi fitur FFT, hingga klasifikasi menggunakan 1D-CNN, dan mengevaluasi performanya baik pada data uji terkontrol maupun pada nada yang diambil dari lagu nyata.

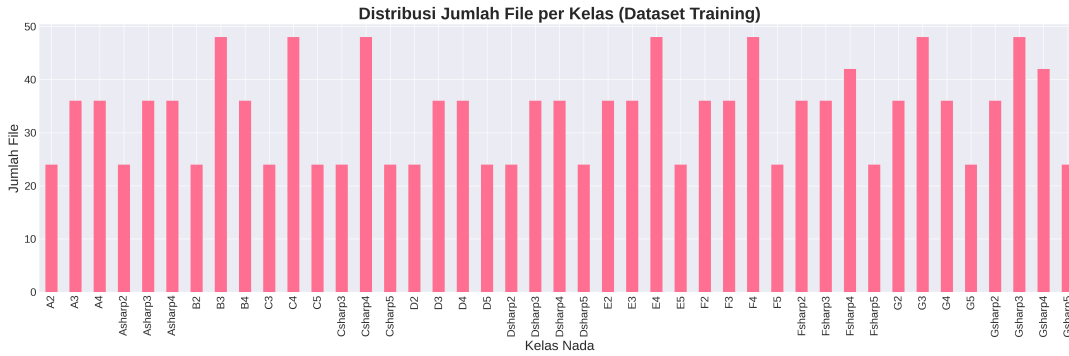
Penelitian ini bertujuan untuk mengimplementasikan teknik ekstraksi fitur FFT pada sinyal nada dasar gitar, menganalisis hasil klasifikasi CNN terhadap tingkat akurasi deteksi nada dasar, serta mengevaluasi kemampuan deteksi nada dasar gitar menggunakan kombinasi FFT dan CNN pada beberapa lagu nyata. Dengan tujuan tersebut, penelitian ini diharapkan dapat memberikan gambaran mengenai efektivitas kombinasi FFT dan CNN dalam mengolah sinyal gitar, baik pada kondisi data yang terkontrol maupun pada kondisi akustik dunia nyata yang lebih kompleks. Secara praktis, hasil penelitian ini juga diharapkan dapat menjadi dasar pengembangan aplikasi bantu penyetelan gitar otomatis yang lebih adaptif terhadap variasi kondisi akustik, sekaligus menjadi rujukan awal bagi penelitian selanjutnya yang ingin mengembangkan sistem deteksi nada pada instrumen musik lain dengan pendekatan serupa.

2 Metode Penelitian

2.1 Dataset

Dataset yang digunakan adalah *Acoustic Guitar Notes Dataset* yang disusun oleh Mohammed Alkoheji dan tersedia secara publik di Kaggle melalui tautan <https://www.kaggle.com/datasets/mohammedalkoheji/guitar-notes-dataset>. Dataset ini terdiri dari 1.428 file audio berformat .wav yang mencakup 43 kelas nada dasar gitar, direkam menggunakan dua jenis gitar (Walden G551E berdawai baja dan Yamaha CM-40 berdawai nilon) dengan jumlah sampel yang seimbang antar keduanya. Rentang nada yang direkam mulai dari D2 (nada terendah) hingga Gsharp5 (nada tertinggi), dengan rentang frekuensi 73,42 Hz hingga 830,61 Hz. Setiap file direkam dengan durasi 2 detik pada frekuensi sampling 44,1 kHz dalam format mono, dan setiap kelas nada memiliki minimal 24 hingga 48 rekaman.

Gambar 1 menunjukkan distribusi jumlah sampel untuk seluruh 43 kelas nada dasar pada dataset ini; sebagian besar kelas memiliki 24–36 sampel, dengan jarak antar nada dasar (dalam satuan semitone) yang relatif konsisten. Karakteristik sinyal tiap kelas nada berbeda pada frekuensi fundamentalnya: kelas nada rendah (misalnya D2, E2) memiliki frekuensi fundamental di bawah 100 Hz dengan periode gelombang yang lebih panjang dan energi harmonik yang tersebar pada beberapa puncak frekuensi rendah, sedangkan kelas nada tinggi (misalnya Fsharp5, Gsharp5) memiliki frekuensi fundamental mendekati 800 Hz dengan periode gelombang yang jauh lebih pendek dan puncak spektral yang lebih terisolasi. Durasi peluruhan (*decay*) sinyal juga bervariasi antar kelas: nada-nada rendah pada umumnya meluruh lebih lambat karena senar yang lebih tebal dan bergetar lebih lama, sedangkan nada-nada tinggi cenderung meluruh lebih cepat. Meskipun demikian, bentuk gelombang (*waveform*) secara umum menyerupai pola petikan gitar akustik yang meluruh secara eksponensial setelah dipetik pada seluruh kelas, sebagaimana ditunjukkan pada contoh sinyal mentah di Gambar 6.



Gambar 1: Distribusi Jumlah Sampel per Kelas Nada pada Acoustic Guitar Notes Dataset

2.2 Ekstraksi Fitur menggunakan FFT

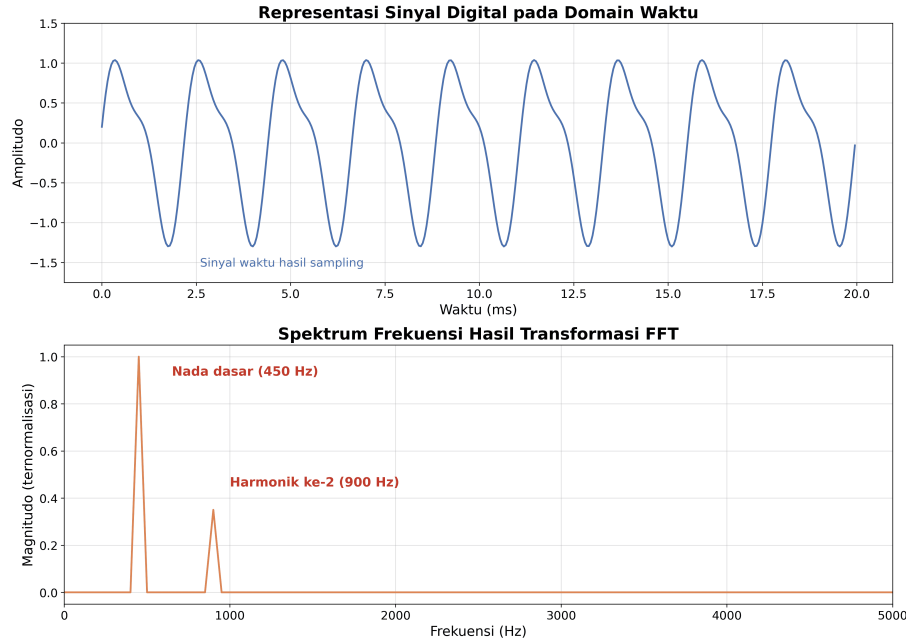
Sinyal suara merupakan gelombang tekanan akustik dalam rentang 20–20.000 Hz yang dapat direpresentasikan sebagai sinyal analog atau digital (Dastres and Soori, 2021). Sinyal digital adalah hasil sampling sinyal analog, direpresentasikan sebagai deretan nilai diskrit. Perubahan amplitudo terhadap waktu (domain waktu) dapat ditransformasikan ke domain frekuensi menggunakan FFT, seperti diilustrasikan pada Gambar 2.

FFT merupakan bentuk efisien dari *Discrete Fourier Transform* (DFT) dengan kompleksitas $\mathcal{O}(N \log N)$ dibandingkan $\mathcal{O}(N^2)$ pada DFT biasa (Shen, 2024). Secara matematis, DFT dirumuskan sebagai:

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j \frac{2\pi}{N} kn}, \quad k = 0, 1, \dots, N-1 \quad (1)$$

di mana $x[n]$ adalah nilai sinyal sampel ke- n , N jumlah total sampel, dan k indeks frekuensi. Algoritma FFT Cooley-Tukey memecah sinyal menjadi bagian genap (E_k) dan ganjil (O_k) untuk mempercepat komputasi:

$$X[k] = E_k + e^{-j \frac{2\pi}{N} k} O_k, \quad k = 0, 1, \dots, \frac{N}{2} - 1 \quad (2)$$



Gambar 2: Representasi Fast Fourier Transform (FFT)

Ketika senar gitar dipetik, getaran menghasilkan sinyal periodik yang dapat dimodelkan sebagai:

$$x(t) = A \sin(2\pi ft + \varphi) \quad (3)$$

dengan A amplitudo, f frekuensi, t waktu, dan φ fase. FFT menguraikan sinyal ini menjadi komponen frekuensi fundamental beserta harmoniknya. Metode ini dipilih karena efisiensi komputasinya yang tinggi serta kemampuannya memberikan representasi frekuensi yang cukup untuk membedakan nada dasar gitar yang bersifat harmonik.

Secara praktis, sinyal audio yang telah melalui tahap praproses (Subbab 2.4) ditransformasikan dari domain waktu ke domain frekuensi menggunakan FFT. Sebelum FFT, diterapkan *Hanning window* untuk mengurangi kebocoran spektrum:

$$y_w(n) = y(n) \cdot w(n) \quad (4)$$

Hasil FFT berupa bilangan kompleks dikonversi ke *magnitude spectrum*. Untuk mengurangi dominasi puncak frekuensi kuat dan mempertahankan detail harmonik, magnitudo dikonversi ke skala logaritmik menggunakan fungsi $\log 1p$, kemudian dinormalisasi dengan z-score. Setiap sampel direpresentasikan sebagai vektor fitur sepanjang 4096 titik frekuensi dengan dimensi input (4096, 1) untuk model CNN.

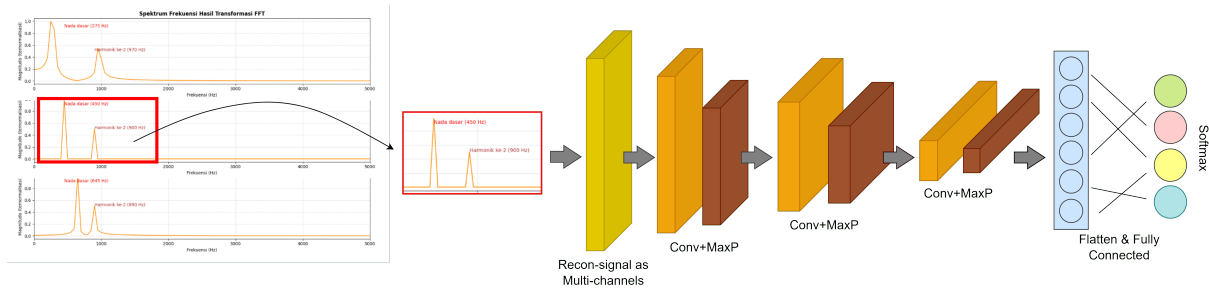
2.3 Klasifikasi Nada Dasar menggunakan CNN

CNN adalah pengembangan *multilayer perceptron* yang dirancang untuk memproses data berpola spasial atau temporal melalui proses konvolusi (Jadhav et al., 2022). Dalam domain audio, CNN mampu mengenali pola temporal dan spektral dari representasi sinyal, terutama ketika data direpresentasikan dalam bentuk domain frekuensi seperti hasil FFT. CNN dipilih dalam penelitian ini karena kemampuannya mengenali pola lokal pada data yang memiliki struktur grid —

termasuk pola harmonik pada spektrum frekuensi — tanpa memerlukan ekstraksi fitur manual tambahan setelah representasi frekuensi diperoleh, serta lebih mampu mempelajari fitur spektral yang kompleks dan bersifat non-linier dibandingkan model klasifikasi konvensional.

Arsitektur CNN pada penelitian ini terdiri dari empat lapisan utama, sebagaimana diilustrasikan pada Gambar 3:

1. **Convolutional Layer:** mengekstraksi pola harmonik dan karakteristik frekuensi dari spektrum FFT menggunakan filter (kernel).
2. **Pooling Layer:** mereduksi dimensi peta fitur untuk efisiensi komputasi dan ketahanan terhadap variasi kecil pada sinyal.
3. **Fully Connected Layer:** mengintegrasikan seluruh fitur menjadi satu vektor representasi global.
4. **Softmax Output Layer:** mengkonversi keluaran menjadi probabilitas untuk setiap kelas nada.



Gambar 3: Arsitektur Convolutional Neural Network (CNN)

Secara konkret, model 1D-CNN yang digunakan dalam penelitian ini terdiri dari:

- Blok 1: Conv1D (32 filter, kernel size 3, stride 2) + MaxPooling1D (pool size 2)
- Blok 2: Conv1D (64 filter, kernel size 3) + MaxPooling1D (pool size 2)
- Flatten + Dense (aktivasi ReLU) + Dropout
- Output: Dense-Softmax (43 kelas)

Model dilatih menggunakan *optimizer* Adam (*learning rate* 0,0001), *batch size* 16, dan maksimum 30 *epoch*. Pembagian data dilakukan dengan proporsi 70% latih (1016 sampel), 15% validasi (218 sampel), dan 15% uji (218 sampel).

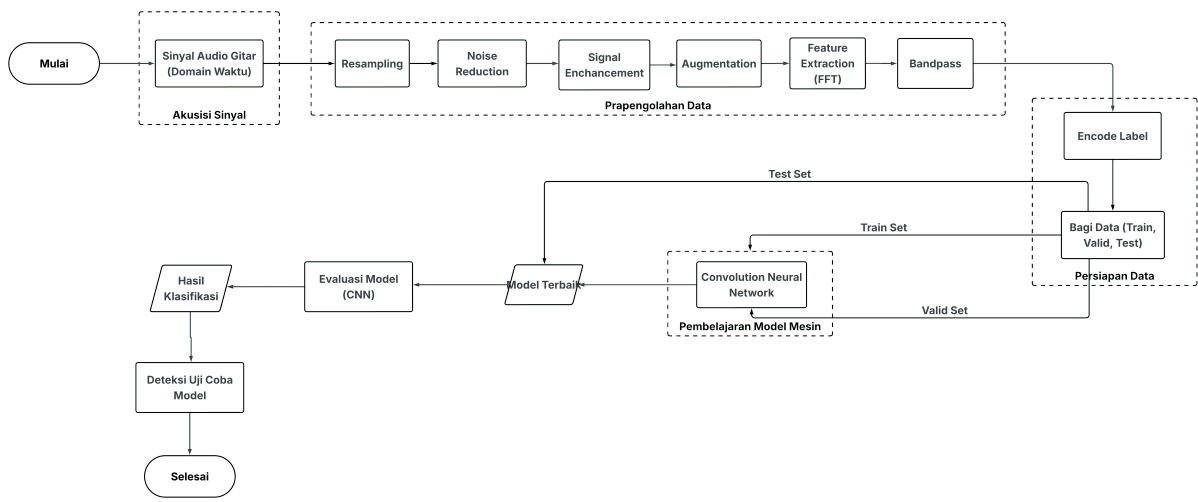
2.4 Prapengolahan Data

Tahap praproses data dilakukan secara berurutan untuk memastikan sinyal audio berada dalam kondisi bersih dan seragam sebelum diekstraksi fiturnya. Tahap pertama adalah *resampling*, yaitu menyeragamkan *sampling rate* seluruh berkas audio ke 44.100 Hz, karena dataset asli memiliki variasi *sampling rate* antar rekaman sehingga perlu disamakan agar representasi frekuensi yang dihasilkan konsisten. Tahap kedua adalah *trim silence*, yaitu memotong segmen hening di awal dan akhir sinyal menggunakan ambang batas (*threshold*) 20 dB, bertujuan menghilangkan bagian sinyal yang tidak mengandung informasi nada agar durasi efektif sinyal lebih fokus pada bagian petikan gitar. Tahap ketiga adalah *bandpass filter* menggunakan filter Butterworth orde 4 dengan rentang frekuensi 70–2000 Hz, yang berfungsi meredam komponen *noise* di luar rentang frekuensi fundamental gitar akustik. Tahap keempat adalah *Wiener filter*, yang digunakan untuk mereduksi *noise* stokastik dengan mengestimasi sinyal bersih $s(t)$ dari sinyal terkontaminasi

$x(t) = s(t) + n(t)$, di mana $n(t)$ merepresentasikan komponen *noise*. Tahap kelima adalah *signal enhancement*, terdiri dari *pre-emphasis* untuk memperkuat energi pada frekuensi tinggi, normalisasi amplitudo ke rentang $[-1, 1]$, serta penyeragaman panjang sinyal menjadi tetap 2 detik menggunakan teknik *padding* (penambahan nilai nol) atau *truncation* (pemotongan) sesuai kebutuhan. Tahap terakhir adalah augmentasi data, yang bertujuan memperkaya variasi data latih dan meningkatkan ketahanan model terhadap variasi sinyal di dunia nyata, meliputi *random gain* ($\alpha \in [0,8; 1,2]$), penambahan *Gaussian noise*, *time shifting* ($\Delta n \in \pm 0,01 \times sr$), *equalization* pada rentang 60–120 Hz dan 1500–3000 Hz secara acak, serta *simple reverb* dengan *delay* 25 ms dan *decay* 0,25. Keluaran dari seluruh tahap praproses ini adalah sinyal audio bersih dan seragam berdurasi 2 detik yang siap diekstraksi fiturnya menggunakan FFT (Subbab 2.4 sebelumnya).

2.5 Alur Penelitian

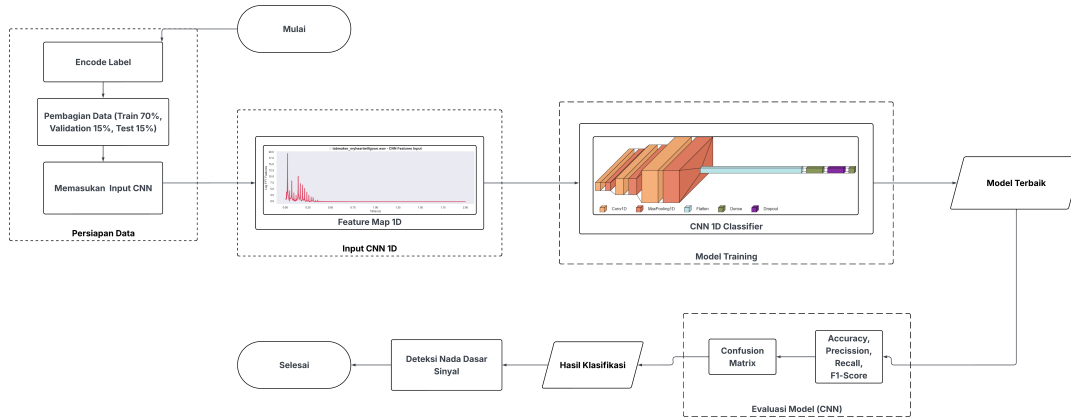
Alur penelitian secara keseluruhan ditunjukkan pada Gambar 4, mulai dari akuisisi sinyal, prapengolahan data, ekstraksi fitur FFT, hingga klasifikasi CNN dan deteksi nada dasar pada lagu uji coba. Setelah data melalui praproses dan ekstraksi fitur, label nada diencode dan data dibagi menjadi data latih, validasi, dan uji. Fitur FFT tersebut kemudian dimasukkan sebagai input model 1D-CNN untuk proses pelatihan. Model dengan performa terbaik pada data validasi dipilih untuk dievaluasi lebih lanjut pada data uji dan pada nada-nada yang diambil dari lagu nyata, sebagaimana diuraikan lebih rinci pada Gambar 5.



Gambar 4: Diagram Alir Penelitian

2.6 Evaluasi

Evaluasi model dilakukan menggunakan empat metrik utama, yaitu *precision*, *recall*, *F1-score*, dan *support*, yang dihitung berdasarkan nilai *True Positive* (TP), *False Positive* (FP), dan *False*



Gambar 5: Blok Diagram Olah Data Setelah Proses FFT

Negative (FN) dari hasil klasifikasi tiap kelas nada:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (6)$$

$$\text{F1-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

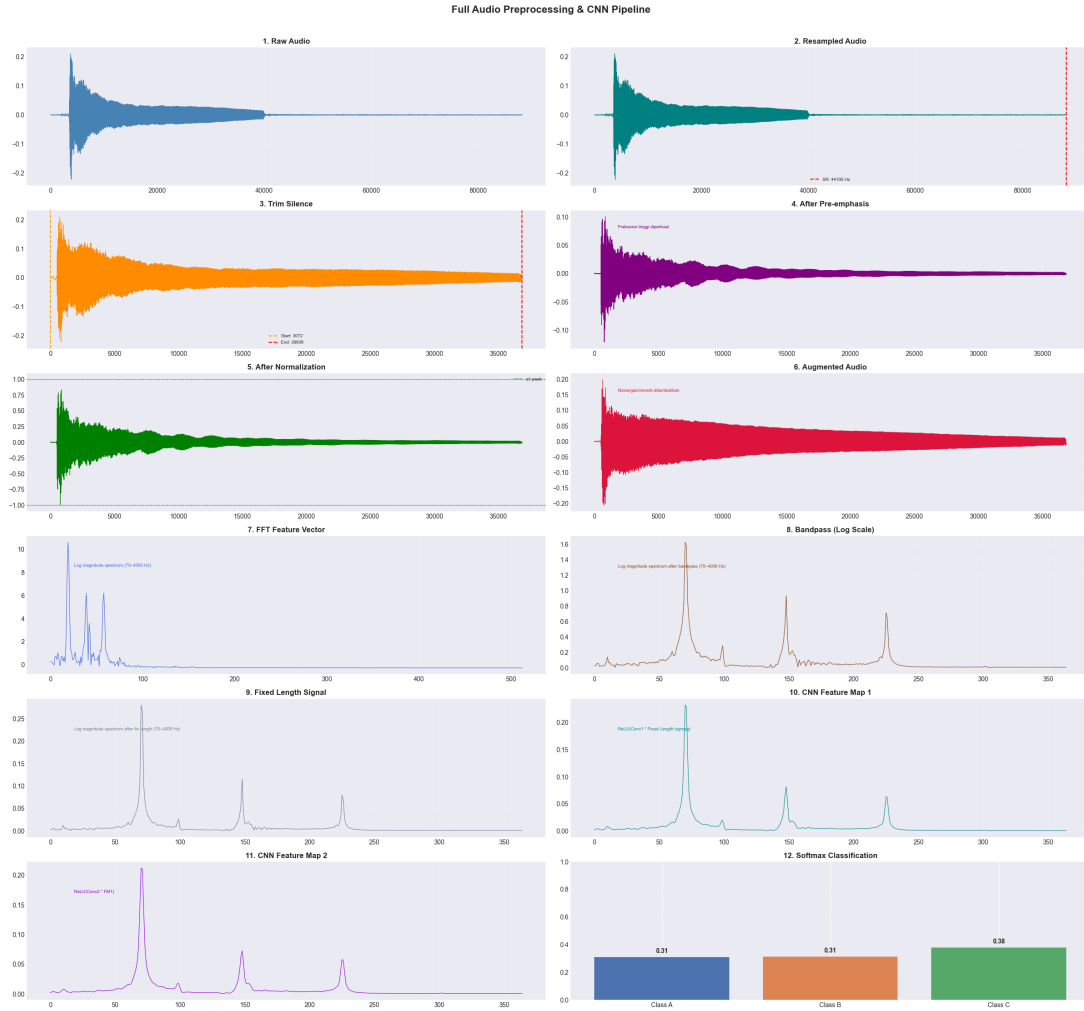
Precision mengukur proporsi prediksi benar terhadap seluruh prediksi suatu kelas, sehingga menunjukkan seberapa tepat model saat menyatakan suatu nada terdeteksi. *Recall* mengukur proporsi prediksi benar terhadap seluruh data aktual suatu kelas, sehingga menunjukkan seberapa baik model menangkap seluruh kemunculan nada tersebut. *F1-score* merupakan rata-rata harmonik dari *precision* dan *recall*, digunakan untuk melihat keseimbangan antara keduanya, terutama penting ketika jumlah sampel antar kelas tidak seimbang. *Support* menunjukkan jumlah sampel aktual pada tiap kelas di data uji.

3 Hasil dan Pembahasan

3.1 Analisis Prapengolahan

Proses prapengolahan menghasilkan sinyal yang lebih bersih dan terfokus. Gambar 6 menampilkan alur pemrosesan lengkap satu sampel sinyal audio sebelum tahap klasifikasi, mulai dari *raw audio*, *resampled audio*, *trim silence*, *pre-emphasis*, normalisasi, augmentasi, FFT *feature vector*, *bandpass* (log scale), *fixed-length signal*, hingga *CNN feature map* dan output *softmax*. Visualisasi ini memperlihatkan perubahan yang terukur pada karakteristik sinyal di setiap tahapan, dan hasil akhirnya adalah representasi fitur yang stabil dan siap diklasifikasikan oleh model CNN.

Bandpass filter 70–2000 Hz berhasil mereduksi komponen *noise* frekuensi rendah (di bawah 70 Hz) berupa *rumble* dan getaran mekanis, serta komponen frekuensi tinggi (di atas 2000 Hz) yang bukan merupakan nada fundamental gitar. Hasilnya terlihat pada panel *bandpass* (log scale) di Gambar 6, di mana puncak frekuensi fundamental tampak lebih tajam dan terisolasi dibandingkan spektrum mentah. Contoh Log FFT nada Gsharp5 menunjukkan puncak fundamental



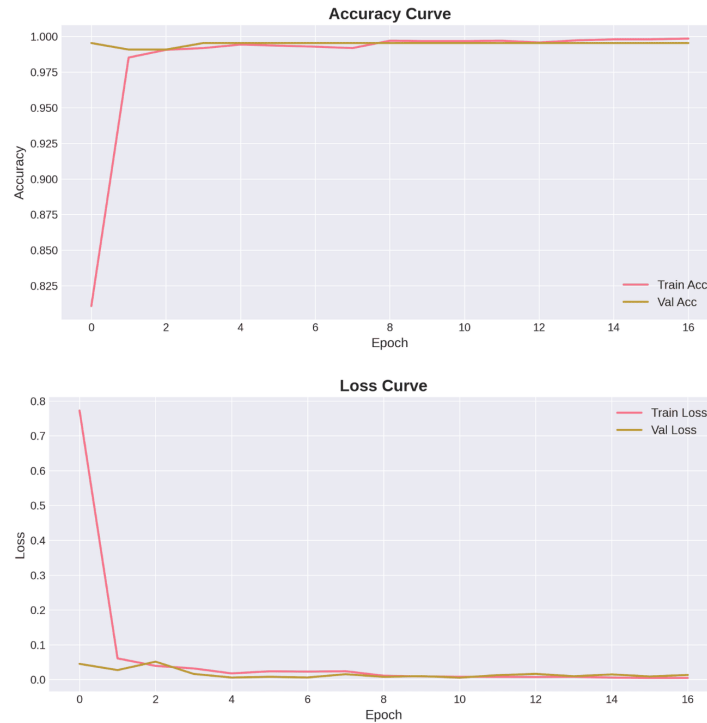
Gambar 6: Hasil Proses Sampel Sinyal Sebelum Klasifikasi

yang konsisten di 829–840 Hz pada tiga file rekaman berbeda, dengan variasi kecil $\approx 10,8$ Hz antar sampel. Variasi ini disebabkan oleh perbedaan tekanan jari pada senar dan kondisi perekaman, namun tetap dalam rentang yang dapat dibedakan oleh model. Panel *CNN feature map* juga memperlihatkan bahwa lapisan konvolusi berhasil mengekstraksi fitur lokal dari spektrum FFT, di mana pola aktivasi yang berbeda terlihat pada tiap kelas nada, mengkonfirmasi bahwa representasi Log FFT mengandung informasi yang cukup untuk membedakan 43 kelas nada dasar

3.2 Hasil Pelatihan Model CNN

Gambar 7 menampilkan kurva akurasi dan *loss* selama proses pelatihan. Kurva akurasi pada data latih meningkat secara konsisten dari sekitar 82% pada *epoch* pertama hingga mendekati 100% pada *epoch* ke-16, sedangkan akurasi validasi mengikuti tren yang sama dan konvergen pada nilai yang berdekatan. Kecilnya selisih antara akurasi latih dan validasi di sepanjang proses pelatihan menunjukkan bahwa model tidak mengalami *overfitting*, artinya model berhasil mempelajari pola umum dari fitur spektral nada gitar, bukan sekadar menghafal data latih.

Kurva *loss* pada keduanya memperlihatkan tren penurunan yang konsisten dari nilai awal sekitar 0,8 hingga mendekati nol pada *epoch* akhir, tanpa lonjakan atau fluktuasi yang signifikan. Hal ini mengindikasikan bahwa proses optimasi menggunakan Adam berjalan stabil dan *learning rate* 0,0001 yang dipilih sudah tepat—tidak terlalu besar sehingga menyebabkan divergensi, dan tidak terlalu kecil sehingga konvergensi tetap tercapai dalam 30 *epoch*. Secara keseluruhan, pola kurva latih dan validasi yang berhimpitan dan stabil mengkonfirmasi bahwa arsitektur 1D-CNN dengan dua blok Conv1D-MaxPooling sudah memadai untuk mengklasifikasikan 43 kelas nada dasar gitar dari representasi Log FFT.



Gambar 7: Plot Kurva Akurasi dan Loss

3.3 Evaluasi pada Data Uji

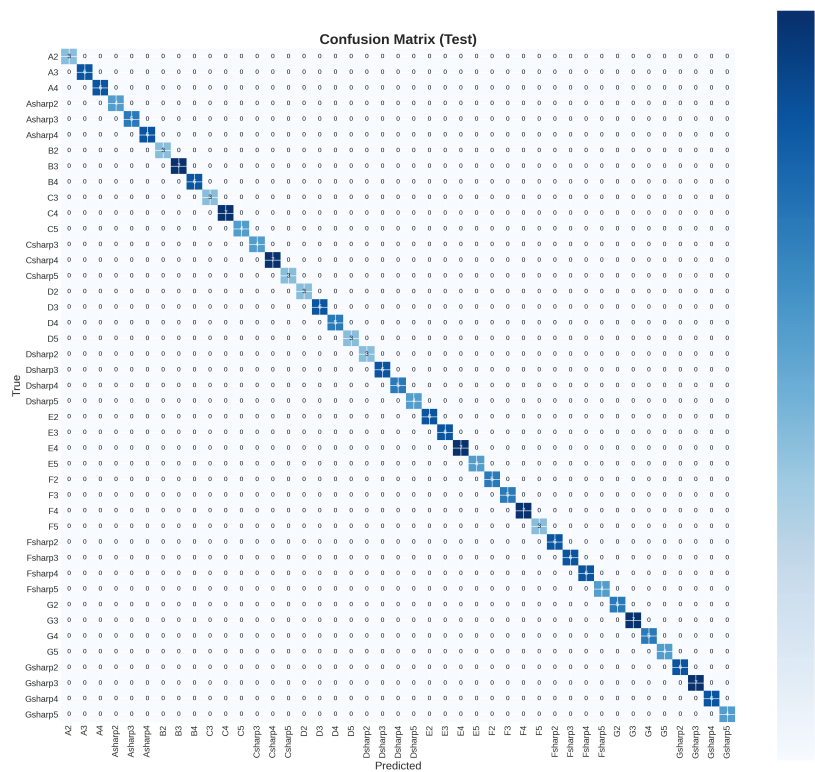
Tabel 1 menampilkan ringkasan *classification report* pada data uji. Hampir seluruh kelas nada mencapai *precision*, *recall*, dan *F1-score* sempurna (1,00), yang berarti model mampu mengenali karakteristik spektral tiap nada dengan sangat baik. Penurunan performa hanya terjadi pada kelas A4 (*F1-score* 0,80) dan Fsharp4 (*F1-score* 0,86). Rendahnya *recall* pada kelas A4 (0,67) mengindikasikan bahwa sebagian sampel A4 justru terklasifikasi sebagai kelas lain, kemungkinan besar kelas dengan frekuensi fundamental yang berdekatan, sehingga spektrum FFT dari kedua nada tersebut memiliki kemiripan yang tinggi dan sulit dibedakan oleh model. Akurasi keseluruhan mencapai **99%** (macro avg = weighted avg = 0,99).

Confusion matrix pada Gambar 8 memperkuat temuan tersebut, di mana kesalahan klasifikasi yang terjadi terkonsentrasi pada kelas-kelas dengan frekuensi fundamental berdekatan, bukan tersebar acak ke seluruh kelas. Hal ini menunjukkan bahwa kesalahan model bersifat sis-

Tabel 1: Ringkasan Classification Report pada Data Uji

Kelas	Precision	Recall	F1-Score	Support
A2	1,00	1,00	1,00	3
A3	1,00	1,00	1,00	6
A4	1,00	0,67	0,80	6
... (39 kelas lain)	1,00	1,00	1,00	—
Fsharp4	0,75	1,00	0,86	6
Accuracy			0,99	218
Macro avg	0,99	0,99	0,99	218
Weighted avg	0,99	0,99	0,99	218

tematis dan berkaitan dengan kemiripan karakteristik spektral, bukan disebabkan oleh ketidakstabilan model secara umum. Nilai pada diagonal utama mendominasi matriks, dengan kesalahan deteksi yang sangat sedikit.



Gambar 8: Confusion Matrix Test Set.

3.4 Deteksi Nada Dasar pada Lagu Nyata

Model diuji pada tiga lagu dari data website (Tabmaker) dan empat percobaan rekaman langsung (lagu Peterpan “Semua Tentang Kita”). Tabel 2 merangkum hasil keseluruhan pada semua data uji lagu nyata.

Tabel 2: Ringkasan Hasil Uji Coba Data Uji pada Lagu Nyata

No	Lagu	Percobaan	Total Nada	Benar	Salah	Akurasi
1	Peterpan	Percobaan 1	7	5	2	0,714
2	Peterpan	Percobaan 2	7	6	1	0,857
3	Peterpan	Percobaan 3	8	7	1	0,875
4	Peterpan	Percobaan 4	7	6	1	0,857
5	Twinkle Twinkle	Single	14	12	2	0,857
6	My Heart Will Go On	Single	21	20	1	0,952
7	Selamat Ulang Tahun	Single	20	19	1	0,950
Rata-rata Akurasi Keseluruhan						0,866

Kesalahan deteksi pada data uji nyata disebabkan perbedaan kondisi rekaman. Data latih berasal dari kondisi terkontrol, sedangkan data uji direkam di lingkungan nyata dengan *noise* latar belakang dan efek akustik (*reverb*, *echo*). Komponen *noise* dan pantulan suara ikut terwakili dalam spektrum FFT, menghasilkan representasi fitur yang berbeda dari pola yang dipelajari model selama pelatihan. Meskipun demikian, akurasi rata-rata 86,6% pada pengujian lagu nyata tetap menunjukkan bahwa model yang dilatih pada data terkontrol mampu melakukan generalisasi cukup baik ke kondisi rekaman yang lebih realistis, walaupun *noise* latar belakang dan efek akustik ruangan tetap menjadi faktor utama yang menurunkan akurasi dibandingkan hasil pada data uji terkontrol (99%).

4 Kesimpulan

Penelitian ini berhasil mengimplementasikan kombinasi FFT dan CNN untuk mendeteksi nada dasar gitar akustik. Tahap praproses data yang komprehensif — meliputi *resampling*, *trim silence*, *bandpass filter* pada rentang 70–2000 Hz, *Wiener filter*, dan *signal enhancement* — terbukti mampu menghasilkan representasi spektrum yang secara efektif mencerminkan karakteristik frekuensi masing-masing nada. Model 1D-CNN berbasis fitur Log FFT yang dikembangkan mencapai akurasi 99% pada data uji (*classification report*), yang membuktikan kemampuan CNN dalam membedakan pola spektral antar nada meskipun memiliki karakteristik frekuensi yang berdekatan. Pada pengujian terhadap lagu nyata, sistem mencapai akurasi rata-rata 86,6%, dengan akurasi pada data website berkisar 85–95% dan data rekaman langsung rata-rata 82%. Hasil ini menunjukkan potensi penerapan sistem pada kondisi nyata, meskipun masih terdapat ruang peningkatan terkait generalisasi model terhadap variasi akustik lingkungan seperti *noise* latar belakang dan efek pantulan suara. Penelitian selanjutnya disarankan untuk memperluas variasi nada dan jenis instrumen yang diuji, mengintegrasikan sistem ke dalam aplikasi *real-time*, serta menguji performa model pada lingkungan akustik yang lebih bervariasi.

Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Program Studi Sarjana Sains Data Universitas Negeri Surabaya serta seluruh pihak yang telah memberikan dukungan dan masukan selama pelaksanaan penelitian ini.

Daftar Pustaka

- Cooley, J. W. and Tukey, J. W. (1965). An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*, 19(90):297–301.
- Dastres, R. and Soori, M. (2021). Artificial neural network systems. *International Journal of Imaging and Robotics*, 21(2):13–25.
- Jadhav, S., Kumar, A., and Singh, R. (2022). Convolutional neural networks for classification of guitar effects. *EURASIP Journal on Audio, Speech, and Music Processing*, 2022(1):1–18.
- Jannereth, C. and Esch, T. (2021). Analyzing timbres of various musical instruments using FFT and spectral analysis. In *Proceedings of the International Conference on Acoustics*, pages 134–141.
- Kow, M. and Kow, J. (2025). A speech enhancement method using fast Fourier transform and convolutional autoencoder. *Applied Acoustics*, 210:109–118.
- Lei, T. (2022). Convolutional neural networks for the classification of guitar effects and extraction of the parameter settings from instrument mixes. *IEEE Signal Processing Letters*, 29:1100–1104.
- Pratama, D., Effendi, R., and Saryadi, B. (2023). Penerapan convolutional neural network untuk melakukan estimasi pitch pada rekaman suara penyanyi. *Jurnal Teknologi Informasi*, 15(2):88–97.
- Shen, H. (2024). A review of fast Fourier transform algorithms and applications. *Digital Signal Processing*, 145:104–115.
- Stefani, L. and Turchet, L. (2025). Automatic guitar tuning using signal processing techniques. *IEEE Transactions on Multimedia*, 27:311–320.
- Wijaya, D. and Adikara, P. (2025). Fourier transformation algorithm for detecting acoustic musical instrument string sounds web-based. volume 16, pages 100–109.