Journal of Information System Research (JOSH)
Volume 6, No. 2, Januari 2025, pp 912−920
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i2.6440

Klasifikasi Irama Murottal Al-Quran Menggunakan Metode CNN dengan Perbandingan Arsitektur ResNet50 dan VGG16

Ilham Rizky Agustin*, Agung Wahana, Aldy Rialdy Atmadja

Fakultas Sains dan Teknologi, Teknik Informatika, UIN Sunan Gunung Djati, Bandung
Jl. A.H. Nasution No.105, Cipadung Wetan, Kec. Cibiru, Kota Bandung, Jawa Barat, Indonesia
Email: [1,*]ilhamrizkyagustin4732@gmail.com, [2]wahana.agung@uinsgd.ac.id, [3]aldyrialdy@uinsgd.ac.id
Email Penulis Korespondensi: ilhamrizkyagustin4732@gmail.com
Submitted: **12/12/2024;** Accepted: **28/12/2024;** Published: **01/01/2025**

Abstrak−Tingkat pemahaman masyarakat Indonesia di bidang murottal Al-Quran saat ini masih tergolong minim. Salah satu faktornya adalah sulitnya membedakan antar irama murottal yang mengharuskan seseorang untuk memiliki keahlian khusus. Di sisi lain, pembelajaran murottal secara tradisional mengharuskan interaksi langsung dengan guru ahli, yang tidak selalu mudah diakses oleh semua kalangan. Tantangan-tantangan ini menekankan pentingnya pengembangan teknologi untuk membantu identifikasi irama murottal. Penelitian ini mengembangkan model klasifikasi irama murottal menggunakan metode Convolutional Neural Network (CNN) berbasis transfer learning dengan dua arsitektur populer, yaitu VGG16 dan ResNet50. Data audio diekstraksi menjadi fitur Short-Time Fourier Transform (STFT) dan Mel-Frequency Cepstral Coefficients (MFCC) untuk dianalisis. Hasil penelitian menunjukkan bahwa arsitektur ResNet50 dengan data ekstraksi MFCC memberikan performa terbaik dengan akurasi pelatihan sebesar 92%, akurasi validasi 85%, dan akurasi pengujian 86%. Selain itu, model ini memiliki nilai precision, recall, dan F1-score masing-masing sebesar 0,87 dan 0,86, yang menunjukkan kemampuan generalisasi yang baik. Sebaliknya, arsitektur VGG16 dengan data ekstraksi STFT dan MFCC menghasilkan akurasi yang lebih rendah dibandingkan ResNet50. Hasil yang diperoleh diharapkan dapat menjadi solusi inovatif dalam pengembangan sistem pembelajaran mandiri berbasis teknologi untuk memahami irama murottal Al-Quran.

Kata Kunci: CNN; Klasifikasi Audio; ResNet50; VGG16; Quran

Abstract−The understanding of murottal Al-Quran among the Indonesian population remains relatively limited. One contributing factor is the difficulty in distinguishing between different murottal rhythms, which requires specialized expertise. Additionally, traditional murottal learning methods necessitate direct interaction with expert teachers, which is not always accessible to everyone. These challenges highlight the importance of developing technology to assist in identifying murottal rhythms. This study developed a murottal rhythm classification model using Convolutional Neural Networks (CNN) with transfer learning, employing two popular architectures: VGG16 and ResNet50. Audio data were processed using Short-Time Fourier Transform (STFT) and Mel-Frequency Cepstral Coefficients (MFCC) feature extraction for analysis.The results showed that the ResNet50 architecture with MFCC-extracted data achieved the best performance, with a training accuracy of 92%, validation accuracy of 85%, and testing accuracy of 86%. Additionally, the model achieved precision, recall, and F1-score values of 0.87 and 0.86, indicating strong generalization capabilities. Conversely, the VGG16 architecture with STFT and MFCC-extracted data demonstrated lower accuracy compared to ResNet50. The findings are expected to provide an innovative solution for developing a self-learning system based on technology to facilitate understanding of murottal rhythms in the Al-Quran.

Keywords: CNN; Audio Classification; ResNet50; VGG16; Quran

1. PENDAHULUAN

Seni membaca Al-Quran telah menjadi tradisi bagi umat islam dari sejak zaman Rasulullah. Rasulullah sendiri mendorong umatnya untuk memperindah suara mereka Ketika melantunkan Al-Quran sebagaimana sabdanya dalam sebuah hadist yang diriwayatkan oleh Abu Daud An-Nasai, Rasulullah bersabda:"Hiasilah Al-Quran dengan suaramu" [1], [2]. Para akademis Islam sepakat bahwa transformasi seni membaca Al-Quran telah terjadi secara turun temurun di daerah Mekah dan Mesir sehingga tercipta berbagai variasi irama yang umum kita dengarkan saat ini [3]. Beberapa irama tersebut antara lain adalah Bayati, Nahawand, dan Jiharkah[4]. Irama-irama tersebut biasa dilantukan dengan berbagai teknik salah satunya adalah teknik murottal, yaitu membaca quran dengan pelan, jelas dan sesuai aturan tajwid[5].

Mempelajari murottal Al-Qur'an secara tradisional memerlukan perhatian tinggi dengan berinteraksi kepada guru yang ahli. Dalam pendekatan ini, seorang harus memperhatikan setiap detail pembacaan yang diajarkan oleh gurunya, mulai dari emosi, nada, hingga pola melodi yang menjadi ciri khas setiap irama. Oleh karena itu proses seseorang untuk mampu memahami dan mengenali setiap irama murottal memerlukan waktu yang tidak sebentar[6]. Di Indonesia, Tingkat pemahaman masyarakat terhadap seni membaca Al-Quran termasuk murottal masih tergolong minim. Suatu penelitian di Pondok Pesantren menyebutkan bahwa kualitas siswanya dalam bidang ini masih tertinggal[7]. Salah satu faktor yang mempengaruhi ketertinggalan ini adalah faktor dalam diri seseorang yang masih merasa kesulitan dalam membedakan jenis-jenis irama murottal[8].

Teknologi saat ini membuka peluang untuk mempermudah dalam proses identifikasi karakteristik suara. Terdapat beberapa penelitian yang telah dilakukan terkait dengan deteksi irama dalam seni membaca Al-Quran salah satunya adalah penelitian yang memanfaatkan berbagai metode untuk klasifikasi maqam atau irama bacaan Al-Quran. Penelitian tersebut memanfaatkan data audio dari 2 qari terkenal, menerapkan beberapa model dan mengahasilkan temuan bahwa model Artificial Neural Network(ANN) dengan fitur spektral dan temporal

Journal of Information System Research (JOSH)
Volume 6, No. 2, Januari 2025, pp 912−920
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i2.6440

mencapai akurasi tertinggi yaitu 95,7% dan F1-Score 0,96[5]. Penelitian yang dilakukan oleh Faisal Omari, memanfaatkan 2 jenis dataset yang lebih besar. Penelitiannya memanfaatkan berbagai ekstraksi fitur dan menghasilkan temuan bahwa model ANN memiliki akurasi tinggi 94,3% dengan ekstraksi fitur MFCC[9]. Penelitian yang menggunakan 3 jenis pengklasifikasi yaitu nearest neighbour, multi-layered perceptron and deep learning terbukti sangat baik dengan akurasi klasifikasi 96% dengan deep learning[10].

Penelitian yang memanfaatkan Convolutional Neural Network(CNN) dan ekstraksi fitur shifted delta cepstral coefficients(SDCC) mendapatkan nilai akurasi 86,05% dalam klasifikasi irama bacaan Al-Quran[11]. Selain itu CNN juga telah digunakan dalam klasifikasi pembaca Al-Quran dengan mencapai nilai akurasi sangat baik yaitu 98%[12]. Dalam penelitian lainnya metode CNN menghasilkan rasio kesalahan kata dan karakter yang minim sebesar 8,3% dan 2,4%[13]. Pengenalan tajwid dalam pengucapan qolqolah dengan metode CNN juga menghasilkan tingkat akurasi tinggi rata-rata di angka 93%[14]. Penggunaan CNN dengan metode ekstraksi MFCC memperoleh performa yang cukup baik dalam klasifikasi huruf hijaiyah[15]. Selain penelitian dengan teknik membaca murottal, penelitian dengan memanfaatkan rekaman gaya mujawwad juga telah dilakukan dengan menggunakan metode Naïve Bayes dengan Tingkat akurasi 56,7%. Rendahnya kinerja model tersebut diakibatkan oleh bias Ketika melakukan ekstraksi[8]. Pemanfaatan Support Vector Machine(SVM) pada data suara yang telah diekstraksi menggunakan MFCC berhasil mengklasifikasi irama Bayati dengan akurasi 94%[16].

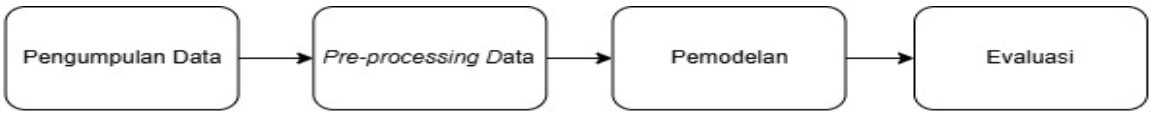
Meskipun telah banyak penelitian terkait klasifikasi suara menggunakan model CNN, belum ada studi yang secara spesifik memanfaatkan arsitektur pre-trained model seperti ResNet50 dan VGG16 untuk klasifikasi irama murottal Al-Quran. Penggunaan transfer learning pada arsitektur CNN dengan memanfaatkan model yang telah dilatih (pre-trained model) dapat menjadi pilihan untuk melakukan pengembangan model klasifikasi suara[17]. Minimnya penelitian ini menjadi tantangan dalam pengembangan sistem otomatis untuk membantu masyarakat Indonesia dalam mengenali dan mempelajari irama murottal secara efektif. Oleh karena itu, penelitian ini berfokus pada eksplorasi dan evaluasi pemanfaatan model pre-trained tersebut dalam mendeteksi irama murottal.

Beberapa arsitektur CNN yang telah digunakan dalam sistem identifikasi suara adalah VGG dan ResNet. Pada penelitian yang dilakukan untuk pengenalan pembicara (Speaker Recognition) kedua arsitektur tersebut diketahui mampu untuk mengenali pembicara dengan Tingkat akurasi tertinggi oleh ResNet dengan nilai 93%[18]. Penggunaan ResNet50 lainnya diterapkan dalam identifikasi gender terhadap data suara menghasilkan akurasi 98,5%[19]. Pada penelitian identifikasi audio terhadap emosi pembicara dengan arsitektur VGG mencapai akurasi tertinggi 87%[20]. Eksplorasi pada penggunaan fitur ekstraksi yang diterapkan pada audio juga dapat dilakukan. Selain dengan MFCC, fitur STFT juga telah diterapkan dalam penelitian klasifikasi genre musik sebagai fitur ekstraksi spektogram dan menghasilkan akurasi 88,9% pada model CNN dengan VGG-16[21]. Kedua ekstraksi fitur tersebut telah berhasil mendeteksi kebohongan berbasis audio dengan akurasi 97% dan 95%[22]. Penelitian lainnya terhadap Fitur STFT diterapkan pada keperluan forensik dalam identifikasi suara dengan hasil akurasi 98%[23].

Arsitektur ResNet50 dan VGG16 dipilih karena berdasarkan penelitian sebelumnya kedua arsitektur tersebut memiliki karakteristik dan performa yang sangat baik dalam masalah klasifikasi. ResNet50 unggul dalam mengenali pola kompleks pada data dengan jaringan yang sangat dalam, sementara VGG16 lebih baik dalam menangkap pola sederhana secara hierarkis. Penelitian ini bertujuan untuk menerapkan dan membandingkan performa model Convolutional Neural Network (CNN) dengan arsitektur ResNet50 dan VGG16 dalam mengklasifikasi irama murottal Al-Quran. Fitur ekstraksi yang digunakan meliputi Mel Frequency Cepstral Coefficients (MFCC) dan Short-Time Fourier Transform (STFT). Kinerja model diukur untuk mengetahui efektivitasnya dalam membedakan jenis-jenis irama murottal, sehingga dapat memberikan gambaran model klasifikasi terbaik dalam deteksi irama murottal Al-Quran.

2. METODOLOGI PENELITIAN

Penelitian ini dilakukan melalui serangkaian tahapan utama yang terstruktur. Tahapan penelitian ini dilakukan untuk memastikan pencapaian hasil dan tujuan penelitian yang optimal. Setiap tahapan dirancang untuk memberikan fokus khusus dari awal penelitian hingga akhir. Dengan pendekatan ini, diharapkan metode yang digunakan dapat menghasilkan model yang efektif dan akurat dalam mengklasifikasi irama murottal. Gambar 1 merupakan ilustrasi alur dari metode penelitian yang dilakukan.



Gambar 1. Alur Metode Penelitian

Berdasarkan Gambar 1, tahapan pertama dimulai dengan pengumpulan dan peninjauan data yang akan digunakan dalam penelitian. Data yang dikumpulkan terdiri dari rekaman murottal Al-Quran dengan berbagai irama yang menjadi fokus klasifikasi. Data yang telah terkumpul kemudian dibagi ke dalam tiga subset, yaitu pelatihan, validasi, dan pengujian. Selanjutnya, data yang telah dibagi memasuki proses pre-processing, yang

Journal of Information System Research (JOSH)
Volume 6, No. 2, Januari 2025, pp 912−920
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i2.6440

mencakup augmentasi data, serta transformasi data menggunakan ekstraksi fitur STFT dan MFCC. Data yang telah diproses kemudian dilatih menggunakan arsitektur CNN VGG16 dan ResNet50 untuk mendeteksi pola pada dataset. Model yang telah selesai dilatih akan dievaluasi menggunakan data uji untuk mengukur kinerjanya dengan perhitungan metrik seperti akurasi, precision, recall, dan F1-Score, dengan tujuan untuk menilai sejauh mana model mampu melakukan klasifikasi terhadap irama murottal Al-Quran secara efektif. Evaluasi ini memberikan gambaran performa model sehingga dapat ditentukan model mana yang memiliki kinerja terbaik dalam klasifikasi irama murottal Al-Quran.

2.1 Pengumpulan Data

Pada tahap ini, dilakukan pengumpulan data audio murottal Al-Quran yang berfokus pada tiga jenis irama, yaitu Bayati, Nahawand, dan Jiharkah. Data audio murottal dikumpulkan dari berbagai sumber untuk memastikan keberagaman dan kualitas yang memadai. Sumber data pada penelitian ini meliputi:

a. Perekaman langsung: Audio direkam langsung dari tiga qori dengan irama yang telah ditentukan.
b. Dataset publik: Dataset rekaman murottal yang tersedia secara online dari platform Figshare.

Seluruh data yang telah dikumpulkan kemudian dikelompokkan secara sistematis menjadi tiga bagian, yaitu 70% data untuk pelatihan, 15% untuk validasi, dan 15% untuk pengujian, ukuran rasio pembagian data ini mengacu pada penelitian sebelumnya [9]. Pembagian data kedalam beberapa set ini dilakukan untuk memastikan model dapat dilatih, divalidasi, dan diuji secara optimal.

2.2 Pre-processing Data

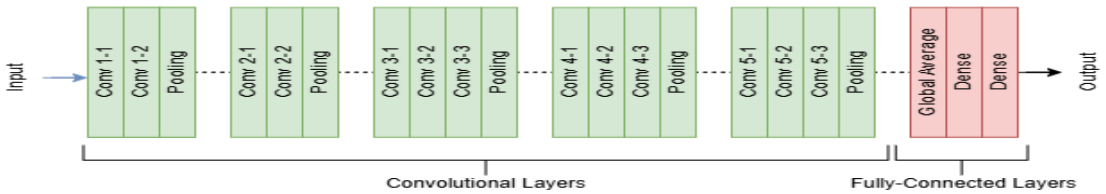
Setelah data dikumpulkan, langkah selanjutnya adalah memproses data untuk mempersiapkannya sebagai masukan model. Proses ini melibatkan augmentasi pada data pelatihan untuk meningkatkan keberagaman dataset, sehingga mampu meningkatkan kemampuan generalisasi model terhadap data baru yang lebih bervariasi. Teknik augmentasi yang digunakan dalam penelitian ini meliputi:

a. Noise injection dilakukan dengan menambahkan noise acak pada data rekaman untuk merepresentasikan variasi lingkungan atau gangguan kecil pada sinyal audio.
b. Time stretching merupakan teknik mengubah kecepatan pemutaran audio sebagai variasi dari tempo bacaan. Pada teknik ini digunakan fungsi dari librosa untuk memperlambat audio dengan faktor 0.8. Durasi audio tetap terjaga dengan memotong bagian yang lebih dari 30 detik.
c. Pitch shifting merupakan proses pengubahan pitch audio untuk mengsimulasikan berbagai karakter suara. Pengubahan pitch ini dilakukan secara acak antara -2 hingga 2 semitone menggunakan fungsi librosa.

Seluruh data yang dikumpulkan ditransformasi ke dalam bentuk spektogram menggunakan fitur ekstraksi Short-Time Fourier Transform (STFT) dan Mel-Frequency Cepstral Coefficients (MFCC). STFT, yang merupakan transformasi Fourier pada bagian lokal sinyal, digunakan untuk menguraikan fungsi dalam domain waktu ke dalam frekuensi penyusunnya, memberikan informasi spektral penuh terkait amplitudo dan frekuensi sepanjang waktu (Fourier Transform)[22]. MFCC merupakan teknik ekstraksi fitur berbasis transformasi cepstral yang mengubah sinyal suara dari domain waktu ke domain frekuensi. Teknik ini mengaplikasikan bank filter Mel pada spektrum daya logaritmik untuk menghasilkan koefisien cepstral yang relevan dengan persepsi pendengaran manusia, memungkinkan analisis pola suara seperti pengenalan irama secara lebih efisien [6], [22]. Setelah transformasi, data disesuaikan dimensi dan ukurannya untuk memastikan konsistensi serta kompatibilitas dengan model yang akan digunakan, sehingga dapat diolah secara optimal.

2.3 Pemodelan

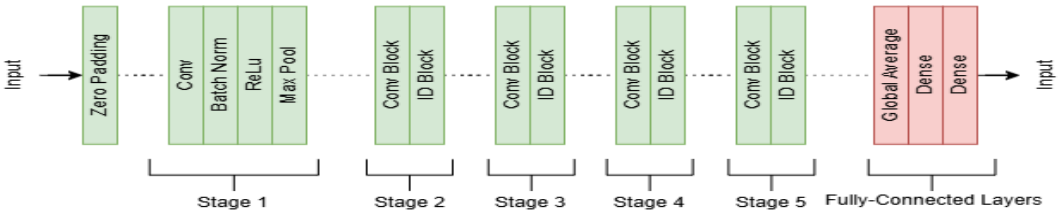
Model yang digunakan pada penelitian ini menerapkan arsitektur CNN, yaitu VGG16 dan ResNet50, untuk membangun model klasifikasi irama murottal. Kedua arsitektur ini dipilih karena memiliki performa yang sangat baik pada berbagai tugas klasifikasi, serta fleksibilitasnya dalam diterapkan pada dataset yang berbeda, termasuk data audio yang telah ditransformasi. Selain itu, kemampuan transfer learning dari model ini memberikan keuntungan dalam meningkatkan akurasi pada model klasifikasi.



Gambar 2. Lapisan Arsitektur VGG16

Gambar 2 merupakan lapisan arsitektur VGG16. Arsitektur ini terdiri dari 13 lapisan konvolusi dengan filter dan pooling, diikuti oleh 3 lapisan fully-connected[24]. Setiap blok konvolusi diakhiri dengan pooling untuk mengurangi dimensi, dan bagian akhirnya menggunakan Global Average Pooling sebelum lapisan Dense untuk klasifikasi. Arsitektur ini memiliki desain yang sederhan, sering digunakan dalam transfer learning, dan dirancang untuk menerima input standar seperti gambar atau representasi lainnya, seperti spektogram.

Journal of Information System Research (JOSH)
Volume 6, No. 2, Januari 2025, pp 912−920
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i2.6440



Gambar 3. Layer Arsitektur ResNet50

Lapisan arsitektur ResNet50 ditampilkan pada gambar 3, jaringan konvolusional dengan fitur utama berupa residual connections yang memungkinkan informasi melewati blok konvolusi melalui koneksi ID Block[24]. Arsitektur ini terdiri dari beberapa stage yang dimulai dengan operasi dasar seperti konvolusi, batch normalization, ReLU, dan max pooling, diikuti oleh blok residual pada setiap stage. Lapisan fully-connected terdiri dari Global Average Pooling untuk merangkum fitur menjadi vektor kecil, diikuti oleh beberapa lapisan Dense, dengan lapisan terakhir yang disesuaikan jumlah kelas dalam dataset.

Kedua model arsitektur dilatih menggunakan data latih dan validasi yang telah diprepreses. Penggunaan data validasi bertujuan untuk memonitor performa model terhadap data yang tidak dilibatkan kedalam pelatihan secara langsung sehingga dapat meminimalisir overviting. Model yang dibangun menerima 2 jenis masukan yaitu data dengan ekstraksi STFT dan MFCC. Seluruh model klasifikasi irama murottal yang digunakan dalam penelitian ini dijelaskan dalam Tabel 1.

Tabel 1. Model Klasifikasi

Model	Jenis Data
CNN VGG16	STFT
CNN ResNet50	STFT
CNN VGG16	MFCC
CNN ResNet50	MFCC

Berdasarkan Tabel 1, setiap arsitektur menerima 2 jenis input data. Jenis data yang menajdi masukan kepada setiap model tersebut antara lain adalah data audio yang diekstraksi melalui STFT dan MFCC. Sehingga model yang digunakan dalam penelitian ini berjumlah 4 model klasifikasi. Seluruh model tersebut akan dilatih dan dievaluasi untuk mengetahui model mana yang terbaik dalam mengklasifikasi irama murottal.

2.4 Evaluasi

Seluruh model yang telah dilatih selanjutnya akan dievaluasi untuk mengukur performa dan akurasinya. Evaluasi dilakukan dengan menggunakan data pengujian yang telah disiapkan dan belum pernah dilihat oleh model. Metode evaluasi dilakukan menggunakan confusion matrix yang merepresentasikan prediksi dan nilai sebenarnya terhadap hasil pengujian model. Tabel 2 merupakan representasi dari confusion matrix terhadap klasifikasi multikelas.

Tabel 2. Confussion Matrix Multikelas

	Prediksi			
	A	**B**	**….**	**N**
A	TP_{AA}	FP_{AB}	…	FP_{AN}
B	FP_{BA}	TP_{BB}	…	FP
…	…	…	…	…
N	FP_{NA}	FP_{NB}	…	TP_{NN}

Berdasarkan confusion matrix dilakukan juga perhitungan metriks lainnya yaitu akurasi, presisi, recall, dan F1-score untuk mendapatkan pemahaman menyeluruh terhadap kemampuan model. Berikut adalah penjelasan dari setiap matriks yang digunakan:

a. Akurasi: Mengukur sejauh mana model dapat memprediksi dengan benar dari total seluruh data.
b. Presisi: Mengukur seberapa banyak prediksi benar untuk suatu kelas dibandingkan dengan seluruh prediksi untuk kelas tersebut.
c. Recall: Mengukur sensitivitas atau kemampuan model untuk mengenali data benar setiap kelas.
d. F1-Score: Mengukur keseimbangan antara presisi dan recall secara rata-rata untuk semua kelas.

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Penelitian ini memerlukan data rekaman murottal Al-Quran yang merepresentasikan tiga irama populer di kalangan qori, yaitu Bayati, Nahawand, dan Jiharkah. Pemilihan ketiga irama ini bertujuan untuk

Journal of Information System Research (JOSH)
Volume 6, No. 2, Januari 2025, pp 912−920
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i2.6440

menyederhanakan analisis dan proses pelatihan model. Adapun data yang digunakan mencakup rekaman langsung dari individu yang memiliki pemahaman mendalam tentang seni baca murottal, serta rekaman yang diambil dari dataset murottal yang tersedia di Figshare.

Data rekaman langsung melibatkan 3 orang qori laki-laki yang masing masing membacakan beberapa ayat Al-Quran dan setiap qori membacakannya dengan 3 irama murottal. Setiap sample rekaman dipotong potong menjadi berdurasi 30 detik sehingga total data dari perekaman langsung berjumlah 78 sample. Sedangkan data yang diperoleh dari dataset murottal yang tersedia di Figshare berjumlah total 246 dari 3 label irama. Seluruh data tersebut dikumpulkan menjadi dataset baru dengan total data 324 sample yang akan digunakan untuk proses pelatihan, validasi, dan pengujian model yang dikembangkan. Tabel 3 merupakan distribusi data untuk setiap kelas irama.

Tabel 3. Distribusi data setiap kelas

Kelas	Jumlah Data
Bayati	108
Jiharkah	108
Nahawand	108

Berdasarkan Tabel 3, distribusi data untuk setiap kelas dilakukan secara merata yaitu 108 data per kelas. Data yang telah dikumpulkan kemudian dibagi menjadi 3 set, yaitu set training, validation, dan testing dengan pembagian 70% untuk data pelatihan, 15% untuk validasi, dan 15% untuk pengujian. Pembagian data ini dilakukan dengan acak menggunakan fungsi dari Scikit-Learn. Setelah pembagian, data data tersebut dimasukan kedalam folder baru berupa Train, Val, dan Test. Setiap folder memiliki sub-folder yang merepresentasikan label kelas sesuai dengan jenis irama murottal yaitu Bayati, Jiharkah, dan Nahawand.

3.2 Pre-processing Data

3.2.1 Augmentasi

Setelah pembagian data dilakukan, data pada folder train yang digunakan untuk pelatihan model diaugmentasi untuk meningkatkan keberagaman dataset. Proses augmentasi ini bertujuan untuk membantu model mempelajari pola yang lebih bervariasi sehingga meningkatkan kemampuan generalisasi terhadap data baru. Dengan demikian, model diharapkan dapat menghasilkan kinerja yang lebih baik dalam mengklasifikasikan irama pada data yang belum pernah dilihat sebelumnya.

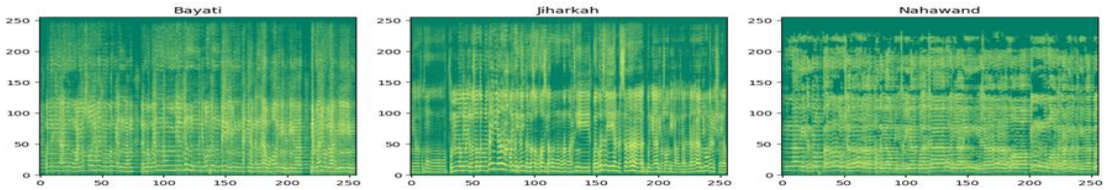
Teknik augmentasi yang diterapkan meliputi noise injection, time stretching, dan pitch shifting, yang diterapkan secara merata pada seluruh data latih. Proses ini menghasilkan data baru yang lebih variatif, memperluas cakupan data pelatihan. Setelah dilakukan augmentasi, jumlah total data pelatihan meningkat secara signifikan dari 225 menjadi 900 data untuk seluruh kelas irama. Peningkatan ini memberikan kontribusi penting dalam menciptakan dataset yang lebih kaya untuk pelatihan model.

3.2.2 Ekstraksi Fitur

Seluruh dataset kemudian ditransformasi dengan menggunakan dua jenis metode ekstraksi. Kedua metode ekstraksi tersebut yaitu Short-Time Fourier Transform (STFT) dan Mel Frequency Cepstral Coefficient (MFCC). Berikut adalah penjelasan hasil dari metode transformasi yang digunakan:

a. STFT

Seluruh data audio diubah ke sample rate 16,000 Hz dan amplitudo dinormalisasi pada rentang -1 hingga 1 untuk menghindari dominasi nilai besar. Sinyal audio kemudian ditransformasi dari domain waktu ke domain frekuensi menggunakan STFT, menghasilkan matriks spektrum yang dikonversi ke skala desibel (dB). Hasil transformasi berupa matriks 2D (frekuensi x waktu) di-resize menjadi ukuran (256, 256) untuk efisiensi komputasi. Matriks yang dihasilkan kemudian dilakukan penambahan dimensi channel, dan dikonversi ke format RGB untuk kompatibilitas dengan arsitektur CNN. Visualisasi hasil STFT ditunjukkan pada Gambar 4.



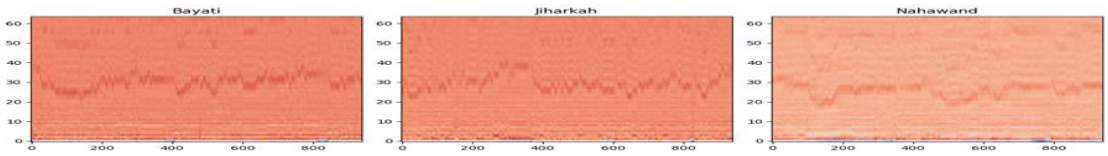
Gambar 4. Visualisasi STFT

b. MFCC

Metode ini dimulai dengan menyelaraskan sample rate menjadi 16,000 Hz. Sinyal audio dibagi menjadi segmen kecil (frame) yang ditransformasi Fourier, menghasilkan matriks frekuensi-waktu. Matriks ini dikonversi ke skala Mel menggunakan filter bank, lalu diubah ke skala logaritmik dan dipadatkan menggunakan DCT untuk menghitung 64 koefisien MFCC per frame, menghasilkan matriks berukuran (64,

Journal of Information System Research (JOSH)
Volume 6, No. 2, Januari 2025, pp 912−920
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i2.6440

938) per audio. Matriks MFCC ditambah dimensi channelnya, dan dikonversi ke format RGB untuk kompatibilitas dengan CNN. Gambar 5 menunjukkan visualisasi matriks MFCC.



Gambar 5. Visualisasi MFCC

3.3 Pemodelan

Terdapat 2 model yang diterapkan dalam penelitian ini yaitu model berbasis CNN dengan arsitektur VGG16 dan model CNN dengan arsitektur ResNet-50. Masing-masing model tersebut dilatih menggunakan 2 inputan yaitu data dengan ekstraksi STFT dan data dengan ekstraksi MFCC. Data yang digunakan selama pelatihan model adalah data latih dan data validasi yang sudah dibagi dari dataset sebelumnya. Data pelatihan digunakan untuk melatih parameter model, sedangkan data validasi digunakan untuk memonitor performa model pada data yang tidak dilibatkan dalam pelatihan langsung, guna mencegah overfitting. Data pengujian yang telah dibagi dari dataset awal sepenuhnya dipisahkan untuk mengukur performa akhir model setelah pelatihan selesai. Berikut adalah penjelasan dari setiap arsitektur model yang digunakan:

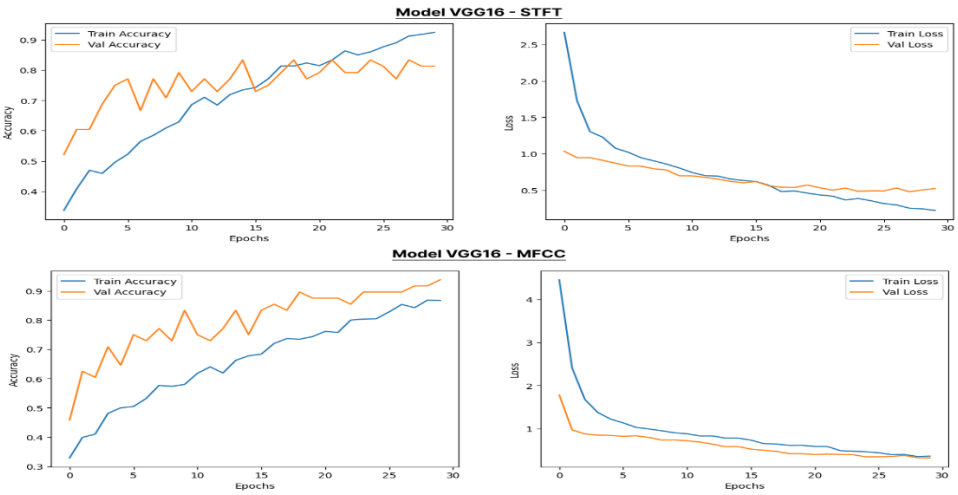
a. VGG16

Metode transfer learning menggunakan model VGG16 yang telah dilatih pada dataset ImageNet. Semua lapisan VGG16 dibekukan agar hanya memanfaatkan fitur pretrained tanpa melatih ulang, untuk efisiensi waktu dan komputasi. Lapisan kustom ditambahkan, termasuk GlobalAveragePooling2D untuk merangkum fitur global, dua lapisan Dense dengan 512 dan 256 unit menggunakan ReLU untuk menangkap pola kompleks, Dropout (0,5) untuk mencegah overfitting, dan lapisan output Dense dengan 3 unit softmax untuk klasifikasi 3 kelas. Model menggunakan optimizer Adam learning rate 0,0005 dan categorical crossentropy untuk klasifikasi multi-kelas, dengan metrik akurasi sebagai evaluasi utama. Total parameter model mencapai 15.109.443. Tabel 4 merupakan lapisan fully connected yang telah disesuaikan.

Tabel 4. Lapisan Custom VGG16

Layer(Type)	Output Shape	Param
…	…	14,714,688
global_average_pooling2d_1	(None, 512)	0
dense_3 (Dense)	(None, 512)	262,656
dropout_2 (Dropout)	(None, 512)	0
dense_4 (Dense)	(None, 256)	131,328
dropout_3 (Dropout)	(None, 256)	0
dense_5 (Dense)	(None, 3)	771

Pelatihan model dilakukan dengan dua data inputan yang berbeda, yaitu data dengan ekstraksi STFT dengan bentuk (256, 256, 3) dan data dengan ekstraksi MFCC dengan bentuk (64, 938, 3). Keduanya dilatih dengan maksimal 30 epoch dan batch size 64. Earlystopping didefinisikan untuk menghentikan pelatihan apabila tidak terjadi perbaikan val_loss selama 5 epoch serta mengurangi learning rate sebesar 50% apabila val_loss tidak membaik selama 3 epoch.



Gambar 6. Grafik Akurasi dan Loss Model VGG16

Journal of Information System Research (JOSH)
Volume 6, No. 2, Januari 2025, pp 912−920
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i2.6440

Berdasarkan grafik akurasi dan loss pada Gambar 6, model VGG16 dengan data MFCC menunjukkan kinerja terbaik dengan akurasi pelatihan mencapai 87% dan akurasi validasi tertinggi 93%, mencerminkan stabilitas dan kemampuan generalisasi yang baik. Penurunan loss yang cepat mengindikasikan model belajar secara signifikan. Sebagai perbandingan, pada model VGG16 dengan data STFT mencapai akurasi pelatihan 92% dan akurasi validasi tertinggi 83%, namun dengan fluktuasi kecil pada validasi, menunjukkan kemampuan generalisasi yang lebih rendah dibandingkan data MFCC.

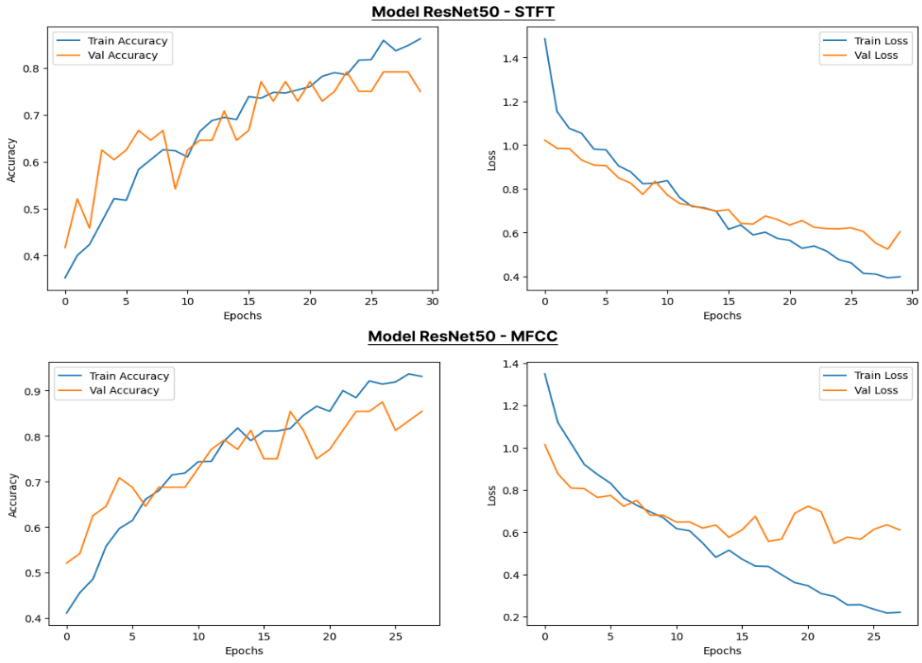
b. ResNet-50

Model ini menggunakan transfer learning dengan ResNet50 yang telah dilatih pada dataset ImageNet. Lapisan pretrained dibekukan untuk mempertahankan parameter awal, sementara lapisan kustom ditambahkan, meliputi GlobalAveragePooling2D, dua Dense yaitu 512 dan 256 unit, Dropout (0,5), dan lapisan output Dense (3 unit, softmax) untuk klasifikasi 3 kelas. Model menggunakan optimizer Adam dengan learning rate 0,0005 dan categorical crossentropy sebagai loss function. Total parameter mencapai 24.768.899. Lapisan custom pada model ini ditampilkan pada Tabel 5.

Tabel 5. Lapisan Custom VGG16

Layer(Type)	Output Shape	Param
…	…	23,587,712
global_average_pooling2d_1	(None, 2048)	0
dense_3 (Dense)	(None, 512)	1,049,088
dropout_2 (Dropout)	(None, 512)	0
dense_4 (Dense)	(None, 256)	131,328
dropout_3 (Dropout)	(None, 256)	0
dense_5 (Dense)	(None, 3)	771

Sama seperti model sebelumnya, melakukan pelatihan dengan 2 jenis data yaitu data dengan ekstraksi STFT bentuk (256, 256, 3) dan data dengan ekstraksi MFCC dengan bentuk (64, 938, 3). Menggunakan maksimal 30 epoch dengan batch size 64. Earlystopping digunakan untuk menghentikan pelatihan dan penyesuaian learningrate apabila tidak ada perbaikan pada val_loss selama beberapa epoch.



Gambar 7. Grafik Akurasi dan Loss Model ResNet50

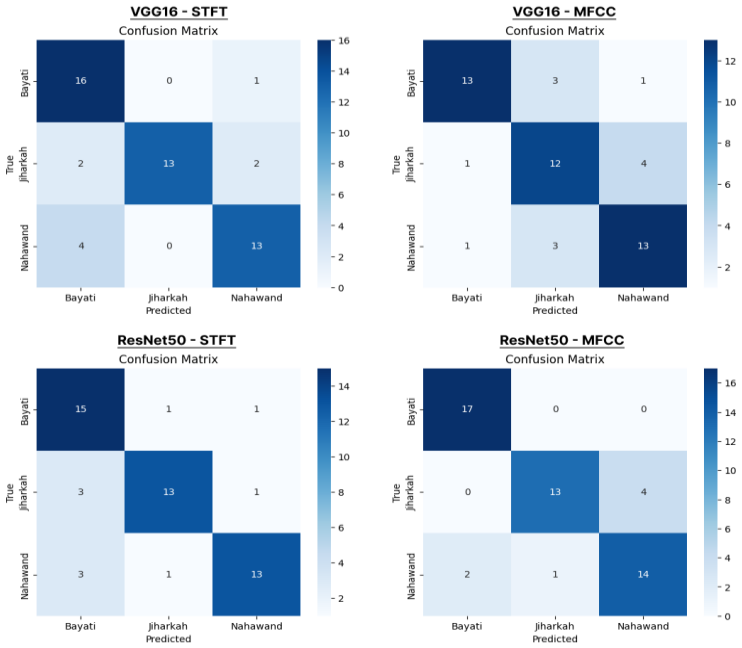
Berdasarkan grafik pada Gambar 7, model ResNet50 dengan data MFCC menunjukkan kinerja terbaik dengan akurasi pelatihan mencapai 92% dan akurasi validasi tertinggi 85%, mencerminkan kemampuan belajar dan generalisasi yang sangat baik. Penurunan loss yang stabil mengindikasikan pembelajaran yang efisien tanpa overfitting. Sebagai perbandingan, pada data STFT (Gambar 8), model mencapai akurasi pelatihan 87% dan akurasi validasi 75%, dengan gap akurasi yang lebih besar dan fluktuasi yang lebih signifikan, menunjukkan performa yang kurang optimal dibandingkan data MFCC.

3.4 Evaluasi

Tahapan evaluasi mencakup pengujian pada setiap model dengan data testing guna mengukur performa model. Model memprediksi data pengujian sehingga menghasilkan probabilitas prediksi dari setiap data irama yang belum

Journal of Information System Research (JOSH)
Volume 6, No. 2, Januari 2025, pp 912−920
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i2.6440

pernah dilihat oleh model. Penggunaan confussion matrix juga diterapkan untuk menghitung jumlah prediksi salah dan benar pada setiap kelas. Gambar 8 adalah hasil evaluasi dari setiap model menggunakan confussion matrix.



Gambar 8. Confussion matrix VGG16 STFT

Pada Gambar 8 diketahui bahwa seluruh model mampu mengklasifikasi dengan cukup baik meskipun masih terdapat kesalahan pada beberapa kelas. Setelah confusion matrix dibuat, metrik evaluasi seperti precision, recall, dan F1-Score dihitung untuk memberikan gambaran yang lebih menyeluruh tentang performa model. Precision mencerminkan akurasi prediksi pada suatu kelas, recall menunjukkan sensitivitas model, dan F1-Score menggabungkan keduanya untuk mengevaluasi keseimbangan. Hasil evaluasi dirangkum dalam Tabel 4, yang membandingkan akurasi pelatihan, validasi, pengujian, serta metrik evaluasi dengan bentuk rata rata dari setiap kelas untuk masing-masing model berbasis VGG16 dan ResNet50 pada input data STFT dan MFCC.

Tabel 4. Perbandingan Akurasi Model

Model	Train	Val	Test	Precission	Recall	F1-Score
VGG16-STFT	92%	83%	82%	0.84	0.82	0.82
VGG16-MFCC	87%	93%	75%	0.75	0.74	0.74
ResNet50-STFT	82%	79%	80%	0.81	0.80	0.80
ResNet50-MFCC	92%	85%	86%	0.86	0.86	0.86

Berdasarkan Gambar 8 dan Tabel 4, performa terbaik ditunjukan pada model ResNet50-MFCC karena keseimbangannya antara akurasi pelatihan, validasi dan pengujian. Akurasi pelatihan yang tinggi di angka 92% diimbangi dengan kinerja yang baik pada validasi (85%) dan pengujian (86%), menunjukkan bahwa model ini tidak mengalami overfitting dan memiliki kemampun generalisasi yang dapat diandalkan. Model mampu memprediksi seluruh data kelas bayati dengan benar. Model ini mencapai nilai precision, recall, serta F1-Score masing-masing berada pada angka 0.86 yang sesuai dengan akurasi prediksi.

Sebagai perbandingan, model VGG16-STFT dan ResNet50-STFT mencapai akurasi prediksi atau pengujian yang sedikit lebih rendah yaitu 82% dan 80%. Kedua model tersebut masih sering salah dalam memprediksi data lain sebagai irama bayati. Sementara model VGG16-MFCC mendapat nilai akurasi prediksi terendah yaitu 75% dengan gap yang cukup jauh dengan nilai akurasi validasinya.

4. KESIMPULAN

Penelitian ini berhasil menerapkan metode CNN untuk klasifikasi irama murottal Al-Quran dengan menggunakan transfer learning untuk membandingkan performa dua arsitektur, yakni VGG16 dan ResNet50, serta mengeksplorasi pengaruh metode ekstraksi fitur STFT dan MFCC. Berdasarkan hasil evaluasi dapat disimpulkan bahwa model ResNet50 dengan ekstraksi fitur MFCC memberikan hasil terbaik dengan akurasi validasi 85% dan akurasi pengujian 86%, serta nilai precision, recall, dan F1-score masing-masing di angka 0.87 dan 0.86. Keseimbangan akurasi tersebut menandakan performa yang stabil dan kemampuan generalisasi yang baik, menunjukkan efektivitas model dalam menghindari overfitting. Penelitian ini membuktikan bahwa transfer learning dapat meningkatkan performa model CNN dalam klasifikasi irama murottal dan memberikan kontribusi

Journal of Information System Research (JOSH)
Volume 6, No. 2, Januari 2025, pp 912−920
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i2.6440

penting bagi pengembangan sistem pembelajaran murottal berbasis teknologi. Penelitian mendatang diharapkan dapat memperluas dataset dengan cakupan yang lebih besar dan variasi irama yang lebih lengkap serta mengeksplorasi arsitektur CNN lainnya untuk meningkatkan kemampuan model dalam mengenali irama murottal Al-Quran di berbagai kondisi nyata.

REFERENCES

[1] C. Supriadi, "Mengenal Ilmu Tadabur Al-Qur'an:(Teori dan Praktek)," ZAD Al-Mufassirin, vol. 4, no. 1, pp. 20–38, 2022, doi: https://doi.org/10.55759/zam.v4i1.34.
[2] W. H. W. Abdullah, W. F. R. W. Mohamad, A. S. M. Razali, and F. I. Zakaria, "Component of Sawt in Hadeeth Texts: Musiqi al-Quran in the Art of Tarannum," AL-TURATH JOURNAL OF AL-QURAN AND AL-SUNNAH, vol. 7, no. 2, pp. 68–77, 2022.
[3] J. Lukita, "Pelestarian dan Pengembangan Nagham Al-Quran: Kajian Resepsi Estetis Al-Quran Di Pondok Pesantren Baitul Qurra Tangerang Selatan," JALSAH: The Journal of Al-Quran and as-Sunnah Studies, vol. 3, no. 2, pp. 1–20, 2023, doi: 10.37252/jqs.v3i2.562.
[4] I. Ilham and K. Kaharuddin, "Pendampingan Program Pondok Pesantren Dalam Penguatan Seni Membaca Al-Qur'an," Jurnal Pema Tarbiyah, vol. 2, no. 1, pp. 10–19, 2023, doi: 10.30829/pema.v2i1.2416.
[5] S. S. Daulay, A. Suciyandhani, S. Sofian, J. Julaiha, and A. Ardiansyah, "Pengenalan Al-Quran," Jurnal Ilmiah Wahana Pendidikan, vol. 9, no. 5, pp. 472–480, 2023, doi: https://doi.org/10.5281/zenodo.7754505.
[6] S. Shahriar and U. Tariq, "Classifying maqams of Qur'anic recitations using deep learning," Ieee Access, vol. 9, pp. 117271–117281, 2021, doi: 10.1109/ACCESS.2021.3098415.
[7] F. Faiza, "Kompetensi Seni Baca Al-Qur'an dalam Meningkatkan Kemampuan Tilawah Santri di Pondok Pesantren an-Najah," Journal of Educational Research, vol. 2, pp. 171–188, Nov. 2023, doi: 10.56436/jer.v2i1.213.
[8] V. Y. Mafula, A. C. Fauzan, and T. R. Fernando, "Identifikasi Irama Tilawah al-Quran dengan Gaya Mujawwad Menggunakan Naive Bayes Classifier," ILKOMNIKA: Journal of Computer Science and Applied Informatics, vol. 4, no. 2, pp. 242–251, 2022, doi: 10.28926/ilkomnika.v4i2.464.
[9] F. Omari, M. Ghantous, and N. Peleg, "Maqam Classification of Quranic Recitations using Deep Learning," figshare, vol. 1, Sep. 2023, doi: https://doi.org/10.6084/m9.figshare.24131781.v1.
[10] A. R. Rababaah, "Intelligent classification model for holy Quran recitation Maqams," Int J Comput Vis Robot, vol. 14, no. 2, pp. 170–190, 2024, doi: 10.1504/IJCVR.2024.136995.
[11] M. A. A. Alaydrus and A. Zahra, "Analysis Of Variation Of Feature Extraction Methods In The Classification Of Al-Qur'an Maqam Using Machine Learning," J Theor Appl Inf Technol, vol. 101, no. 21, 2023.
[12] M. Lataifeh, A. Elnagar, I. Shahin, and A. B. Nassif, "Arabic audio clips: Identification and discrimination of authentic cantillations from imitations," Neurocomputing, vol. 418, pp. 162–177, 2020, doi: 10.1016/j.neucom.2020.07.099.
[13] A. Al Harere and K. Al Jallad, "Quran recitation recognition using end-to-end deep learning," arXiv preprint, vol. 1, 2023, doi: /10.48550/arXiv.2305.07034.
[14] D. M. Omran, A. H. Kandil, A. ElBialy, S. Samy, and S. Fawzy, "CNN for speech recognition case study: Recitation Rules of the holy Quran," MSA Engineering Journal, vol. 2, no. 4, pp. 1–12, 2023, doi: 10.21608/msaeng.2023.225120.1335.
[15] A. M. H. Azis et al., "Automatic Detection of Hijaiyah Letters Pronunciation using Convolutional Neural Network Algorithm," Jurnal Online Informatika, vol. 7, no. 1, pp. 123–131, 2022, doi: doi.org/10.15575/join.v7i1.882.
[16] T. Hidayat, "Klasifikasi jenis irama qiro'ah menggunakan metode Mel frequency cepstral coefficients dan algoritma Support vector machine," Skripsi, Teknik Informatika, UIN Sunan Gunung Djati, Bandung, Indonesia, 2023.
[17] E. Tsalera, A. Papadakis, and M. Samarakou, "Comparison of Pre-Trained CNNs for Audio Classification Using Transfer Learning," Journal of Sensor and Actuator Networks, vol. 10, no. 4, 2021, doi: 10.3390/jsan10040072.
[18] M. Jakubec, E. Lieskovska, and R. Jarina, "Speaker recognition with resnet and vgg networks," in 2021 31st International Conference Radioelektronika (RADIOELEKTRONIKA), IEEE, 2021, pp. 1–5. doi: 10.1109/RADIOELEKTRONIKA52220.2021.9420202.
[19] A. A. Alnuaim et al., "Speaker gender recognition based on deep neural networks and ResNet50," Wirel Commun Mob Comput, vol. 2022, no. 1, p. 4444388, 2022, doi: 10.1155/2022/4444388.
[20] S. Hamsa, I. Shahin, Y. Iraqi, E. Damiani, A. B. Nassif, and N. Werghi, "Speaker identification from emotional and noisy speech using learned voice segregation and speech VGG," Expert Syst Appl, vol. 224, p. 119871, 2023, doi: 10.1016/j.eswa.2023.119871.
[21] T. Toshniwal, P. Tandon, and P. Nithyakani, "Music Genre Recognition Using Short Time Fourier Tranform And CNN," in 2022 International Conference on Computer Communication and Informatics (ICCCI), IEEE, 2022, pp. 1–4. doi: 10.1109/ICCCI54379.2022.9740939.
[22] D. Kusumawati, A. A. Ilham, A. Achmad, and I. Nurtanio, "Performance Analysis of Feature Mel Frequency Cepstral Coefficient and Short Time Fourier Transform Input for Lie Detection using Convolutional Neural Network," JOIV: International Journal on Informatics Visualization, vol. 8, no. 1, pp. 279–288, 2024.
[23] S. Saleem, A. Dilawari, and U. G. Khan, "Spoofed voice detection using dense features of stft and mdct spectrograms," in 2021 International Conference on Artificial Intelligence (ICAI), IEEE, 2021, pp. 56–61. doi: 10.1109/ICAI52203.2021.9445259.
[24] M. Aatila, M. Lachgar, H. Hrimech, and A. Kartit, "Diabetic retinopathy classification using ResNet50 and VGG-16 pretrained networks," International Journal of Computer Engineering and Data Science (IJCEDS), vol. 1, no. 1, pp. 1–7, 2021.