JITSI Vol 1 No. 1 (2026)
Journal of Information Technology and System Integration
https://journal2.ummat.ac.id/index.php/jitsi

Klasifikasi Deepfake Voice Menggunakan Ekstraksi Fitur MFCC dan Random Forest

Rama Nanda[1*], Fachrur Rahman[2], Cahaya Tamila[3], Dyah Isty Sucitra[4]
[1,2]Sistem dan Teknologi Informasi, Universitas Muhammadiyah Mataram
email: nandarama374@gmail.com[1*]

Abstrak:
Perkembangan teknologi kecerdasan buatan telah mendorong munculnya teknologi deepfake voice yang mampu menghasilkan suara sintetis menyerupai suara manusia asli. Meskipun memiliki berbagai manfaat, teknologi ini juga berpotensi disalahgunakan untuk pemalsuan identitas dan penyebaran informasi palsu sehingga diperlukan metode deteksi yang efektif. Penelitian ini bertujuan untuk mendeteksi audio deepfake menggunakan metode Mel Frequency Cepstral Coefficients (MFCC) dan algoritma Random Forest. Dataset yang digunakan adalah Fake-or-Real (FoR) Dataset yang terdiri atas audio REAL dan FAKE. Sebanyak 4.500 sampel dipilih secara seimbang, namun satu file gagal diproses sehingga digunakan 4.499 sampel. Proses penelitian meliputi preprocessing, ekstraksi 13 fitur MFCC, pembagian data menggunakan metode train-test split dengan rasio 80:20, serta klasifikasi menggunakan Random Forest dengan parameter n_estimators = 100. Hasil pengujian menunjukkan bahwa model memperoleh akurasi sebesar 97%, dengan nilai precision, recall, dan F1-score yang seimbang pada kedua kelas. Hasil tersebut menunjukkan bahwa kombinasi MFCC dan Random Forest mampu membedakan audio REAL dan FAKE secara efektif sehingga berpotensi diterapkan pada bidang keamanan digital dan forensik audio.
Kata Kunci: audio deepfake; deteksi suara; MFCC; random forest; voice deepfake

Abstract :
The advancement of artificial intelligence technology has led to the emergence of *deepfake voice* technology capable of generating synthetic speech that closely resembles human voices. Although it offers various benefits, this technology can also be misused for identity fraud and misinformation, creating the need for effective detection methods. This study aims to detect deepfake audio using Mel Frequency Cepstral Coefficients (MFCC) and the Random Forest algorithm. The dataset used is the Fake-or-Real (FoR) Dataset, consisting of REAL and FAKE audio samples. A total of 4,500 audio samples were selected, but one file failed to be processed, resulting in 4,499 samples being used. The research process included preprocessing, extraction of 13 MFCC features, dataset splitting using an 80:20 train-test ratio, and classification using a Random Forest model with *n_estimators* = 100. Experimental results showed that the proposed model achieved an accuracy of 97%, with balanced precision, recall, and F1-score values for both classes. These findings indicate that the combination of MFCC and Random Forest can effectively distinguish REAL and FAKE audio and has potential applications in digital security and audio forensics.
Keywords: audio deepfake; MFCC; random forest; voice deepfake; voice detection

PENDAHULUAN

Perkembangan teknologi kecerdasan buatan (*Artificial Intelligence* atau AI) dalam beberapa tahun terakhir telah membawa perubahan yang signifikan pada berbagai bidang, termasuk pengolahan sinyal suara dan teknologi sintesis ucapan. Kemajuan model generatif memungkinkan pembuatan suara sintetis yang semakin menyerupai suara manusia asli sehingga sulit dibedakan secara langsung oleh pendengaran manusia. Salah satu bentuk implementasi teknologi tersebut adalah *deepfake voice*, yaitu teknologi yang mampu menghasilkan atau meniru suara seseorang berdasarkan sampel suara yang tersedia. Teknologi ini banyak dimanfaatkan dalam bidang hiburan, pendidikan, layanan pelanggan digital, pembuatan konten multimedia, hingga pengembangan asisten virtual karena mampu menghasilkan suara yang alami dan realistis [1]–[3].

Di samping berbagai manfaat yang ditawarkan, perkembangan teknologi *deepfake voice* juga menimbulkan berbagai risiko keamanan yang perlu mendapat perhatian serius. Suara sintetis dapat digunakan untuk melakukan pemalsuan identitas, manipulasi komunikasi, penyebaran informasi palsu, hingga penipuan berbasis suara yang memanfaatkan sistem autentikasi biometrik. Seiring meningkatnya kualitas suara hasil sintesis, kemampuan manusia untuk membedakan suara asli dan suara palsu menjadi semakin terbatas. Kondisi tersebut menjadikan audio *deepfake* sebagai salah satu ancaman yang penting dalam bidang keamanan siber, forensik digital, dan perlindungan identitas digital [1]–[4].

Meningkatnya kasus penyalahgunaan teknologi *deepfake voice* mendorong banyak peneliti untuk mengembangkan sistem deteksi yang mampu mengidentifikasi perbedaan antara suara asli dan suara sintetis. Berbagai pendekatan telah diusulkan, mulai dari analisis karakteristik spektral suara, ekstraksi fitur akustik, hingga penerapan algoritma *machine learning* dan *deep learning* [2], [3], [4]. Penelitian oleh Frank dan Schönherr [5] melalui *WaveFake Dataset* menunjukkan bahwa karakteristik suara sintetis masih memiliki pola tertentu yang dapat dimanfaatkan untuk proses deteksi. Selain itu, pengembangan *ASVspoof Challenge* turut memberikan kontribusi penting dalam menyediakan kerangka evaluasi yang terstandarisasi untuk penelitian deteksi suara hasil manipulasi dan *spoofing* [6].

Salah satu teknik ekstraksi fitur yang banyak digunakan dalam pengolahan sinyal suara adalah *Mel Frequency Cepstral Coefficients* (MFCC). Metode ini diperkenalkan untuk merepresentasikan karakteristik suara berdasarkan skala Mel yang menyesuaikan dengan cara manusia memersepsikan frekuensi suara [7]. MFCC bekerja dengan mengekstraksi

informasi spektral penting dari sinyal audio sehingga mampu menggambarkan karakteristik unik suatu suara dalam bentuk vektor numerik. Karena kemampuannya dalam merepresentasikan informasi frekuensi secara efektif, MFCC telah banyak digunakan pada berbagai aplikasi seperti pengenalan pembicara, pengenalan ucapan, klasifikasi emosi, serta deteksi audio sintetis [10], [11], [12], [15]. Beberapa penelitian menunjukkan bahwa fitur MFCC mampu memberikan performa yang baik dalam membedakan suara asli dan suara hasil sintesis karena dapat menangkap perbedaan pola spektral yang tidak mudah dikenali secara langsung oleh pendengaran manusia [11], [12], [18].

Selain pemilihan fitur, keberhasilan sistem deteksi juga dipengaruhi oleh algoritma klasifikasi yang digunakan. Salah satu algoritma yang banyak diterapkan pada berbagai permasalahan klasifikasi adalah *Random Forest*. Algoritma yang diperkenalkan oleh Breiman [8] ini merupakan metode *ensemble learning* yang membangun sejumlah *decision tree* secara acak dan menggabungkan hasil prediksi setiap pohon untuk menghasilkan keputusan akhir. Pendekatan tersebut terbukti mampu meningkatkan stabilitas model dan mengurangi risiko *overfitting* dibandingkan penggunaan satu *decision tree* tunggal [8], [9]. Selain memiliki tingkat akurasi yang baik, Random Forest juga relatif mudah diimplementasikan, mampu menangani data dengan jumlah fitur yang cukup banyak, serta tidak memerlukan proses pelatihan yang kompleks [17], [19]. Karakteristik tersebut menjadikan Random Forest sebagai salah satu algoritma yang sesuai untuk penelitian deteksi audio *deepfake* berbasis fitur MFCC.

Meskipun berbagai penelitian terkini menunjukkan performa yang baik dalam mendeteksi audio *deepfake*, sebagian besar pendekatan yang digunakan masih berbasis *deep learning* dengan arsitektur yang kompleks dan membutuhkan sumber daya komputasi yang besar [2], [3], [13], [14], [20]. Penggunaan model yang kompleks sering kali memerlukan waktu pelatihan yang lebih lama, kapasitas memori yang tinggi, serta jumlah data yang besar. Kondisi tersebut dapat menjadi kendala pada lingkungan penelitian maupun implementasi yang memiliki keterbatasan perangkat keras. Oleh karena itu, diperlukan pendekatan yang lebih sederhana, efisien, dan mudah diterapkan tanpa mengurangi kemampuan model dalam melakukan klasifikasi audio *deepfake* secara akurat.

Berdasarkan permasalahan tersebut, penelitian ini mengusulkan penerapan metode MFCC sebagai teknik ekstraksi fitur dan algoritma Random Forest sebagai metode klasifikasi untuk mendeteksi audio *deepfake*. Dataset yang digunakan adalah *Fake-or-Real Dataset* (FoR Dataset) yang diperoleh dari Kaggle dan terdiri atas kategori audio REAL dan FAKE [21]. Dari dataset tersebut dipilih sebanyak 4.500 sampel audio yang terdiri atas 2.250 audio REAL dan 2.250 audio FAKE. Tahapan penelitian meliputi pengumpulan dan pemahaman dataset, *preprocessing* data, ekstraksi fitur MFCC, pembagian dataset menggunakan *train-test split*, pelatihan model Random Forest, serta evaluasi model menggunakan metrik *accuracy*, *precision*, *recall*, *F1-score*, dan *confusion matrix*.

Hasil penelitian menunjukkan bahwa kombinasi MFCC dan Random Forest mampu menghasilkan tingkat akurasi sebesar 97% dalam membedakan audio REAL dan audio FAKE. Hasil tersebut menunjukkan bahwa pendekatan berbasis *machine learning* yang relatif sederhana tetap mampu memberikan performa klasifikasi yang tinggi sehingga berpotensi menjadi solusi yang efektif dan efisien dalam mendukung deteksi audio *deepfake* pada berbagai lingkungan komputasi.

TINJAUAN PUSTAKA

Deepfake voice merupakan teknologi sintesis suara yang memanfaatkan kecerdasan buatan untuk menghasilkan atau meniru suara seseorang dengan tingkat kemiripan yang tinggi. Perkembangan model generatif modern menyebabkan kualitas suara sintetis semakin sulit dibedakan dari suara asli, sehingga menimbulkan tantangan baru dalam bidang keamanan siber, autentikasi biometrik, dan forensik digital [1]–[4]. Berbagai penelitian menunjukkan bahwa audio deepfake dapat digunakan untuk pemalsuan identitas, manipulasi komunikasi, serta penyebaran informasi palsu yang berpotensi merugikan individu maupun organisasi [2], [3]. Oleh karena itu, pengembangan metode deteksi audio deepfake menjadi salah satu topik penelitian yang terus berkembang dalam beberapa tahun terakhir [4]–[6].

Mel Frequency Cepstral Coefficients (MFCC) merupakan salah satu metode ekstraksi fitur yang paling banyak digunakan dalam pengolahan sinyal suara. Metode ini diperkenalkan oleh Davis dan Mermelstein [10] untuk merepresentasikan karakteristik frekuensi suara berdasarkan skala Mel yang menyesuaikan persepsi pendengaran manusia. Proses pembentukan MFCC meliputi pre-emphasis, framing, windowing, Fast Fourier Transform (FFT), Mel Filter Bank, dan Discrete Cosine Transform (DCT) sehingga menghasilkan koefisien yang mampu menggambarkan karakteristik spektral suatu sinyal suara [7], [10]. MFCC telah digunakan secara luas pada berbagai aplikasi seperti pengenalan ucapan, identifikasi pembicara, klasifikasi emosi, hingga deteksi audio sintetis [11], [12], [15]. Penelitian sebelumnya menunjukkan bahwa fitur MFCC mampu menangkap perbedaan karakteristik spektral antara suara asli dan suara hasil sintesis sehingga efektif digunakan sebagai representasi fitur pada sistem deteksi deepfake voice [11], [12], [18].

Random Forest merupakan algoritma machine learning berbasis ensemble learning yang diperkenalkan oleh Breiman [8]. Algoritma ini bekerja dengan membangun sejumlah decision tree secara acak dan menggabungkan hasil prediksi masing-masing pohon menggunakan mekanisme voting untuk menentukan kelas akhir. Pendekatan tersebut memungkinkan Random Forest menghasilkan performa klasifikasi yang stabil serta mengurangi risiko overfitting dibandingkan penggunaan decision tree tunggal [8], [9]. Selain memiliki tingkat akurasi yang tinggi, Random Forest juga

mampu menangani data berdimensi menengah, tidak sensitif terhadap noise, serta relatif mudah diimplementasikan [17], [19]. Oleh karena itu, algoritma ini banyak digunakan dalam berbagai permasalahan klasifikasi, termasuk klasifikasi berbasis fitur audio dan pengolahan sinyal suara [13], [17].

Dataset yang digunakan dalam penelitian ini adalah Fake-or-Real Dataset (FoR Dataset) yang tersedia melalui platform Kaggle [21]. Dataset tersebut terdiri atas audio suara asli (REAL) dan audio hasil sintesis (FAKE) yang dikumpulkan untuk mendukung penelitian deteksi audio deepfake. FoR Dataset banyak digunakan sebagai dataset benchmark karena menyediakan jumlah data yang besar dengan variasi karakteristik suara yang cukup beragam [21]. Pada penelitian ini digunakan sebanyak 4.500 sampel audio yang terdiri atas 2.250 audio REAL dan 2.250 audio FAKE. Setelah proses ekstraksi fitur dilakukan, satu file audio tidak dapat diproses sehingga total data yang digunakan dalam penelitian berjumlah 4.499 sampel audio.

Beberapa penelitian terdahulu telah menerapkan berbagai metode untuk mendeteksi audio deepfake. Altalahin dkk. [11] menggunakan fitur MFCC untuk mendeteksi audio deepfake dan menunjukkan bahwa representasi fitur berbasis MFCC mampu membedakan karakteristik suara asli dan sintetis secara efektif. Sivaramakrishnan dkk. [12] juga menerapkan MFCC pada klasifikasi audio deepfake dan memperoleh hasil yang baik dalam proses identifikasi suara manipulasi. Di sisi lain, Ali dkk. [13] menerapkan pendekatan ensemble learning untuk deteksi voice deepfake dan menunjukkan peningkatan performa klasifikasi dibandingkan metode tunggal. Sementara itu, berbagai penelitian lain lebih banyak menggunakan pendekatan deep learning seperti CNN, ResNet, maupun RawNet2 yang mampu menghasilkan akurasi tinggi, tetapi memerlukan sumber daya komputasi yang besar [14], [18], [20].

Berdasarkan penelitian-penelitian tersebut, masih terdapat kebutuhan akan metode deteksi audio deepfake yang lebih sederhana dan efisien tanpa mengurangi performa klasifikasi. Oleh karena itu, penelitian ini menggabungkan metode MFCC sebagai teknik ekstraksi fitur dan Random Forest sebagai algoritma klasifikasi untuk menghasilkan sistem deteksi deepfake voice yang ringan, mudah diimplementasikan, serta tetap mampu memberikan tingkat akurasi yang tinggi.

METODE

Penelitian ini menggunakan pendekatan Machine Learning untuk mendeteksi suara deepfake berdasarkan karakteristik sinyal audio. Proses penelitian terdiri atas enam tahapan utama, yaitu pengumpulan dataset, preprocessing data, ekstraksi fitur MFCC, pembagian dataset, pelatihan model Random Forest, serta evaluasi model. Tahapan penelitian ditunjukkan pada Gambar 3.1.

Gambar 3.1.Tahapan Penelitian

3.1 Pengumpulan Dataset

Penelitian ini menggunakan *Fake-or-Real Dataset (FoR Dataset)* yang diperoleh dari Kaggle [8]. Dataset tersebut terdiri atas dua kategori utama, yaitu audio asli (*REAL*) dan audio hasil sintesis (*FAKE*). Secara keseluruhan, dataset memiliki 69.316 file audio yang terbagi ke dalam data pelatihan (*training*), validasi (*validation*), dan pengujian (*testing*). Pada penelitian ini digunakan 4.500 sampel audio yang terdiri atas 2.250 audio REAL dan 2.250 audio FAKE. Pemilihan sampel dilakukan secara acak dengan jumlah yang seimbang pada masing-masing kelas sehingga dapat mengurangi potensi ketidakseimbangan data saat proses pelatihan model. Analisis karakteristik dataset menunjukkan bahwa audio REAL memiliki rata-rata durasi 5,39 detik dengan durasi minimum 0,90 detik dan maksimum 10,66 detik. Sementara itu, audio FAKE memiliki rata-rata durasi 2,31 detik dengan durasi minimum 0,57 detik dan maksimum 6,12 detik. Frekuensi sampling rata-rata pada audio REAL sebesar 20.984 Hz, sedangkan pada audio FAKE sebesar 18.635 Hz. Perbedaan karakteristik durasi dan frekuensi sampling tersebut menunjukkan adanya variasi sinyal audio yang dapat dimanfaatkan dalam proses ekstraksi fitur dan klasifikasi.

Tabel 3.1. Karakteristik Dataset Audio

Karakteristik	REAL	FAKE

Jumlah Sampel Dianalisis	500	499
Rata-rata Durasi (s)	5,39	2,31
Durasi Minimum (s)	0,90	0,57
Durasi Maksimum (s)	10,66	6,12
Rata-rata Sampling Rate (Hz)	20.984	18.635

3.2 Preprocessing Data

Tahap preprocessing dilakukan untuk mempersiapkan data audio sebelum proses ekstraksi fitur. Adapun tahapan yang dilakukan adalah sebagai berikut:

1. **Pembacaan File Audio**
   File audio dibaca menggunakan pustaka Librosa dengan mempertahankan frekuensi sampling asli (*sampling rate*) menggunakan parameter sr=None.
2. **Validasi Data Audio**
   Setiap file audio diperiksa untuk memastikan dapat diproses dengan baik. File yang mengalami kerusakan atau gagal dibaca akan dikeluarkan dari proses penelitian.
3. **Pelabelan Data**
   Audio REAL diberikan label 0, sedangkan audio FAKE diberikan label 1. Pelabelan dilakukan untuk membedakan kelas pada proses klasifikasi.
4. **Framing Sinyal Audio**
   Pada saat proses ekstraksi MFCC, sinyal audio secara otomatis dibagi menjadi beberapa frame pendek agar karakteristik frekuensi suara dapat dianalisis secara lebih detail.

Hasil dari tahap preprocessing berupa data audio yang valid dan telah memiliki label kelas sehingga siap digunakan pada proses ekstraksi fitur MFCC. Dari 4.500 file audio yang dipilih, sebanyak 4.499 file audio berhasil diproses dan digunakan dalam penelitian, sedangkan 1 file audio gagal diproses karena tidak dapat dibaca oleh sistem.

3.3 Ekstraksi Fitur MFCC

Ekstraksi fitur dilakukan menggunakan metode *Mel Frequency Cepstral Coefficients* (MFCC). Metode ini dipilih karena mampu merepresentasikan karakteristik frekuensi suara manusia secara efektif dan banyak digunakan pada penelitian pengolahan sinyal suara [4], [5]. Proses ekstraksi MFCC diawali dengan pembacaan sinyal audio yang telah melalui tahap preprocessing. Selanjutnya, sinyal audio dibagi menjadi beberapa segmen pendek (*framing*) agar karakteristik frekuensi pada setiap bagian sinyal dapat dianalisis dengan lebih baik. Setiap frame kemudian ditransformasikan dari domain waktu ke domain frekuensi menggunakan *Fast Fourier Transform* (FFT). Hasil transformasi frekuensi selanjutnya dipetakan ke skala Mel melalui *Mel Filter Bank* untuk menyesuaikan karakteristik pendengaran manusia. Setelah itu dilakukan perhitungan koefisien MFCC yang merepresentasikan informasi spektral sinyal audio.

Pada penelitian ini digunakan 13 koefisien MFCC untuk setiap file audio. Nilai MFCC yang diperoleh dari setiap frame kemudian dirata-ratakan sehingga menghasilkan satu vektor fitur untuk setiap sampel audio. Perhitungan rata-rata MFCC dapat dinyatakan pada Persamaan (1).

$$MFCC_n = \frac{1}{T}\sum_{t=1}^{T} C_n \quad (1)$$

dengan $MFCC_n$ merupakan nilai rata-rata koefisien $MFCC_n$ ke $n, c_n(t)$merupakan nilai koefisien $MFCC$ pada frame ke-t dan T merupakan jumlah frame audio. Hasil proses ekstraksi menghasilkan 13 fitur $MFCC$ untuk setiap sampel audio. Selanjutnya fitur-fitur tersebut digunakan sebagai data masukan (*input features*) pada algoritma Random Forest untuk proses klasifikasi audio REAL dan FAKE.

3.4 Training Random Forest

Setelah proses ekstraksi fitur selesai, diperoleh sebanyak 4.499 sampel audio yang terdiri atas fitur MFCC dan label kelas. Data tersebut kemudian dibagi menjadi data pelatihan dan data pengujian menggunakan metode *train-test split* dengan rasio 80:20. Hasil pembagian menghasilkan 3.600 data pelatihan dan 900 data pengujian. Pembagian dilakukan secara acak menggunakan parameter *random_state* = 42 untuk menjaga konsistensi hasil eksperimen. Proses klasifikasi dilakukan menggunakan algoritma Random Forest. Algoritma ini merupakan metode *ensemble learning* yang membangun sejumlah pohon keputusan (*decision tree*) dan menggabungkan hasil prediksi dari setiap pohon untuk menghasilkan keputusan akhir yang lebih akurat [6].

Pada penelitian ini, model Random Forest dibangun menggunakan parameter awal *n_estimators* = 100 yang menunjukkan jumlah pohon keputusan yang digunakan, serta *random_state* = 42 untuk memastikan proses pelatihan dapat direproduksi. Model kemudian dilatih menggunakan data pelatihan yang telah diekstraksi fiturnya menggunakan MFCC. Prediksi akhir pada Random Forest diperoleh melalui mekanisme *majority voting* yang dapat dinyatakan pada Persamaan (2).

$$y = mode\left(h_1(x), h_2(x), \ldots., h_n(x)\right) \quad (2)$$

dengan y merupakan hasil prediksi akhir, $h_i(x)$ merupakan hasil prediksi dari pohon keputusan ke-i, dan n merupakan jumlah pohon keputusan yang digunakan dalam model Random Forest. Setelah proses pelatihan selesai, model digunakan untuk memprediksi kelas audio pada data pengujian ke dalam kategori REAL atau FAKE.

3.5 Evaluasi Model dan Hyperparameter Tuning

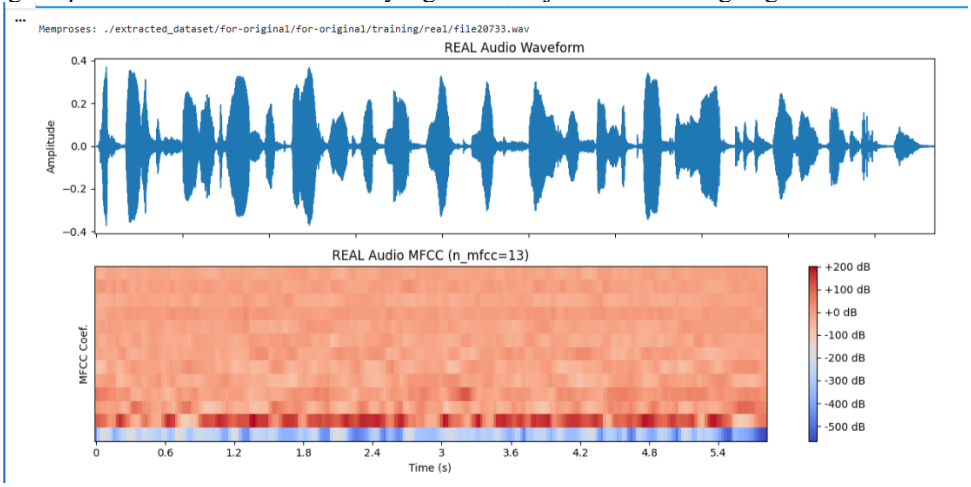
Evaluasi model dilakukan menggunakan data pengujian yang berjumlah 900 sampel audio. Kinerja model diukur menggunakan beberapa metrik evaluasi, yaitu *accuracy*, *precision*, *recall*, *F1-score*, dan *confusion matrix*. Metrik tersebut digunakan untuk mengetahui kemampuan model dalam membedakan audio REAL dan audio FAKE secara tepat. Nilai *accuracy* digunakan untuk mengukur persentase prediksi yang benar terhadap seluruh data pengujian. *Precision* menunjukkan tingkat ketepatan model dalam mengidentifikasi suatu kelas, sedangkan *recall* menunjukkan kemampuan model dalam menemukan seluruh data yang termasuk dalam kelas tersebut. Sementara itu, *F1-score* merupakan rata-rata harmonis antara *precision* dan *recall* sehingga dapat memberikan gambaran performa model secara lebih seimbang [9], [10].

Selain menggunakan metrik evaluasi, penelitian ini juga memanfaatkan *confusion matrix* untuk melihat jumlah prediksi yang benar maupun salah pada masing-masing kelas. Berdasarkan hasil pengujian, model Random Forest berhasil memperoleh akurasi sebesar 97% dengan nilai *precision*, *recall*, dan *F1-score* yang seimbang pada kedua kelas. Hasil tersebut menunjukkan bahwa model mampu membedakan audio REAL dan FAKE dengan tingkat ketepatan yang tinggi.

HASIL DAN PEMBAHASAN

4.1 Hasil Ekstraksi Fitur MFCC

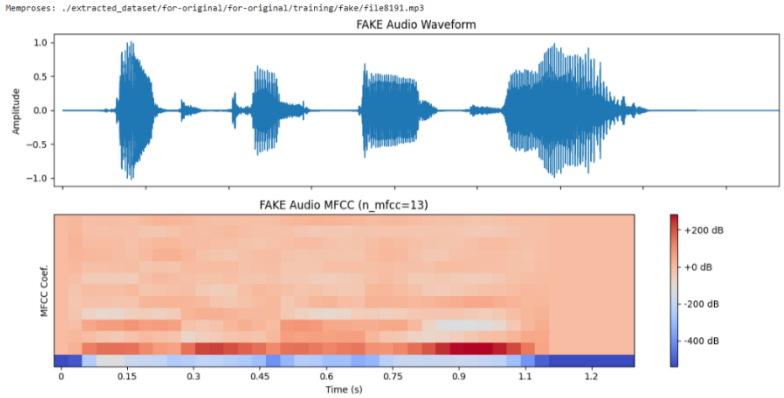
Pada penelitian ini, proses ekstraksi fitur dilakukan menggunakan metode Mel Frequency Cepstral Coefficients (MFCC) dengan jumlah koefisien sebanyak 13 untuk setiap file audio. Metode MFCC dipilih karena mampu merepresentasikan karakteristik frekuensi suara berdasarkan sistem pendengaran manusia sehingga banyak digunakan pada berbagai penelitian pengenalan suara dan klasifikasi audio [7], [10]. Proses ekstraksi diawali dengan pembacaan sinyal audio yang telah melalui tahap preprocessing. Selanjutnya sinyal suara dibagi menjadi beberapa frame dan dilakukan perhitungan MFCC pada masing-masing frame. Nilai koefisien MFCC yang diperoleh kemudian dirata-ratakan sehingga menghasilkan satu vektor fitur yang terdiri atas 13 nilai numerik untuk setiap sampel audio. Vektor fitur tersebut digunakan sebagai representasi karakteristik suara yang akan menjadi masukan bagi algoritma Random Forest.



Gambar 4.1. Waveform dan hasil ekstraksi MFCC pada audio REAL.

Berdasarkan Gambar 4.1, waveform audio REAL menunjukkan pola amplitudo yang berubah secara dinamis sepanjang durasi sinyal. Perubahan amplitudo tersebut menunjukkan adanya variasi energi suara yang merupakan karakteristik alami dari ucapan manusia. Selain itu, hasil ekstraksi MFCC memperlihatkan distribusi koefisien yang relatif

beragam pada setiap frame waktu. Variasi tersebut menunjukkan bahwa suara asli memiliki karakteristik spektral yang kompleks dan berubah-ubah sesuai dengan intonasi, tekanan suara, serta pengucapan setiap kata.



Gambar 4.2. Waveform dan hasil ekstraksi MFCC pada audio FAKE.

Sementara itu, pada Gambar 4.2 terlihat bahwa waveform audio FAKE memiliki pola amplitudo yang lebih teratur dibandingkan audio REAL. Beberapa bagian sinyal menunjukkan bentuk gelombang yang cenderung berulang dan memiliki distribusi energi yang lebih seragam. Hasil ekstraksi MFCC pada audio FAKE juga memperlihatkan pola koefisien yang berbeda dibandingkan audio REAL. Perbedaan ini menunjukkan bahwa suara sintetis yang dihasilkan oleh sistem kecerdasan buatan masih memiliki karakteristik spektral tertentu yang dapat dibedakan dari suara manusia asli. Perbedaan pola antara audio REAL dan audio FAKE menunjukkan bahwa fitur MFCC mampu menangkap informasi penting yang berkaitan dengan struktur frekuensi suara. Informasi tersebut menjadi dasar bagi model Random Forest untuk mempelajari pola dan melakukan klasifikasi audio ke dalam kategori REAL maupun FAKE. Secara keseluruhan, proses ekstraksi fitur berhasil menghasilkan 13 fitur MFCC untuk setiap sampel audio. Dari total 4.500 audio yang dipilih, terdapat satu file yang gagal diproses sehingga jumlah data yang berhasil diekstraksi menjadi 4.499 sampel. Seluruh data hasil ekstraksi kemudian digunakan pada tahap pelatihan dan pengujian model klasifikasi.

4.2 Hasil Klasifikasi Random Forest

Setelah proses ekstraksi fitur MFCC selesai dilakukan, data hasil ekstraksi selanjutnya digunakan sebagai masukan pada algoritma Random Forest untuk melakukan klasifikasi audio ke dalam dua kategori, yaitu REAL dan FAKE. Dari total 4.500 sampel audio yang dipilih pada tahap pengumpulan data, terdapat satu file yang gagal diproses pada saat ekstraksi fitur sehingga jumlah data yang digunakan pada penelitian ini menjadi 4.499 sampel. Data tersebut kemudian dibagi menjadi data pelatihan dan data pengujian menggunakan metode *train-test split* dengan rasio 80:20. Sebanyak 3.600 data digunakan sebagai data pelatihan (*training data*), sedangkan 900 data digunakan sebagai data pengujian (*testing data*). Pembagian data dilakukan secara acak menggunakan parameter *random_state* = 42 untuk menjaga konsistensi hasil eksperimen.

Model Random Forest dibangun menggunakan parameter *n_estimators* = 100 yang berarti model membentuk 100 pohon keputusan (*decision tree*) selama proses pelatihan. Setiap pohon keputusan dilatih menggunakan subset data yang berbeda sehingga mampu meningkatkan kemampuan generalisasi model dan mengurangi risiko terjadinya *overfitting*. Distribusi data pada tahap pengujian terdiri atas 446 sampel audio REAL dan 454 sampel audio FAKE. Komposisi data yang relatif seimbang tersebut membantu proses evaluasi model menjadi lebih objektif karena tidak didominasi oleh salah satu kelas tertentu.

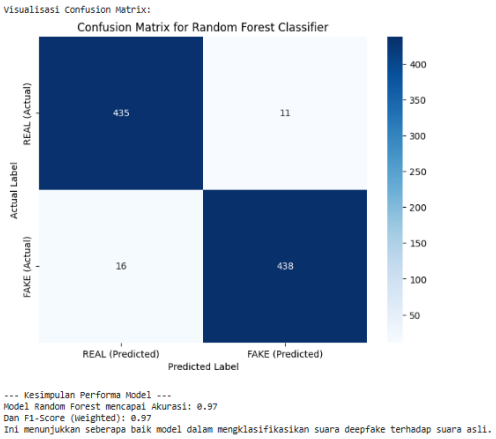
Tabel 4.1 Hasil Evaluasi Model Random Forest

Kelas	Precision	Recall	F1-Score	Support
REAL	0,96	0,98	0,97	446
FAKE	0,98	0,96	0,97	454
Accuracy	-	-	0,97	900
Macro Average	0,97	0,97	0,97	900
Weighted Average	0,97	0,97	0,97	900

Berdasarkan hasil evaluasi pada Tabel 4.1, model Random Forest berhasil memperoleh nilai akurasi sebesar 97% pada data pengujian. Nilai tersebut menunjukkan bahwa sebagian besar data audio berhasil diklasifikasikan dengan benar ke dalam kategori REAL maupun FAKE. Selain nilai akurasi yang tinggi, metrik evaluasi lainnya seperti precision, recall, dan F1-score juga menunjukkan performa yang sangat baik pada kedua kelas. Pada kelas REAL diperoleh nilai precision sebesar 96%, recall sebesar 98%, dan F1-score sebesar 97%. Nilai recall yang tinggi menunjukkan bahwa sebagian besar audio REAL berhasil dikenali dengan benar oleh model. Dengan kata lain, hanya sebagian kecil audio REAL yang salah diklasifikasikan sebagai audio FAKE. Sementara itu, pada kelas FAKE diperoleh nilai precision sebesar 98%, recall

sebesar 96%, dan F1-score sebesar 97%. Nilai precision yang tinggi menunjukkan bahwa sebagian besar audio yang diprediksi sebagai FAKE memang benar merupakan audio sintetis. Hasil ini mengindikasikan bahwa model mampu mengenali karakteristik audio deepfake dengan baik tanpa menghasilkan banyak prediksi salah.
Untuk melihat distribusi prediksi model secara lebih rinci, dilakukan visualisasi menggunakan *confusion matrix* seperti yang ditunjukkan pada Gambar 4.3.

Gambar 4.3. Confusion Matrix hasil klasifikasi Random Forest.



Berdasarkan Gambar 4.4, dari 446 data audio REAL yang diuji, sebanyak 435 data berhasil diklasifikasikan dengan benar sebagai REAL, sedangkan 11 data salah diklasifikasikan sebagai FAKE. Pada kelas FAKE, dari total 454 data yang diuji, sebanyak 438 data berhasil dikenali dengan benar sebagai FAKE, sedangkan 16 data salah diklasifikasikan sebagai REAL. Secara keseluruhan, hanya terdapat 27 kesalahan klasifikasi dari total 900 data pengujian. Jumlah kesalahan yang relatif kecil tersebut menunjukkan bahwa model mampu mempelajari pola karakteristik suara dengan baik berdasarkan fitur MFCC yang digunakan. Tingginya jumlah prediksi yang benar pada kedua kelas juga menunjukkan bahwa model memiliki kemampuan yang seimbang dalam mendeteksi audio REAL maupun audio FAKE. Hasil pengujian ini membuktikan bahwa kombinasi metode MFCC dan Random Forest mampu menghasilkan performa klasifikasi yang tinggi untuk mendeteksi deepfake voice. Selain memperoleh akurasi yang baik, model juga menunjukkan kestabilan performa yang ditunjukkan oleh nilai precision, recall, dan F1-score yang relatif seimbang pada kedua kelas. Kondisi tersebut mengindikasikan bahwa model tidak mengalami bias terhadap salah satu kelas dan mampu melakukan klasifikasi secara konsisten pada data yang digunakan dalam penelitian.

4.3 Pembahasan

Hasil pengujian menunjukkan bahwa kombinasi metode Mel Frequency Cepstral Coefficients (MFCC) dan Random Forest mampu mengklasifikasikan audio REAL dan FAKE dengan akurasi sebesar 97%. Selain memperoleh nilai akurasi yang tinggi, model juga menghasilkan nilai precision, recall, dan F1-score yang relatif seimbang pada kedua kelas. Hal tersebut menunjukkan bahwa model memiliki kemampuan klasifikasi yang baik tanpa kecenderungan memihak salah satu kelas, sehingga mampu mengenali karakteristik audio asli maupun audio sintetis secara konsisten. Performa model dipengaruhi oleh kemampuan metode MFCC dalam merepresentasikan karakteristik spektral sinyal suara. Melalui proses framing, transformasi ke domain frekuensi, pembentukan Mel Filter Bank, hingga perhitungan koefisien cepstral, MFCC menghasilkan representasi numerik yang mampu mempertahankan informasi penting dari sinyal audio. Representasi tersebut memudahkan model Random Forest dalam mempelajari pola perbedaan antara audio REAL dan audio FAKE sehingga proses klasifikasi dapat dilakukan secara efektif. Berdasarkan hasil visualisasi waveform dan MFCC pada Gambar 4.2 dan Gambar 4.3, terlihat adanya perbedaan karakteristik antara kedua jenis audio. Audio REAL menunjukkan variasi amplitudo dan distribusi energi yang lebih dinamis sebagai akibat dari proses produksi suara manusia secara alami. Sebaliknya, audio FAKE memiliki pola amplitudo dan distribusi energi yang cenderung lebih seragam karena dihasilkan melalui proses sintesis menggunakan kecerdasan buatan. Perbedaan karakteristik tersebut menghasilkan nilai koefisien MFCC yang berbeda sehingga dapat dimanfaatkan sebagai dasar dalam proses klasifikasi.
Selain metode ekstraksi fitur, kualitas dataset juga memberikan kontribusi terhadap performa model. Penelitian ini menggunakan 4.499 sampel audio yang berhasil diekstraksi, terdiri atas audio REAL dan FAKE dengan jumlah yang relatif seimbang. Distribusi data yang seimbang membantu model mempelajari karakteristik masing-masing kelas secara lebih optimal sehingga mampu mengurangi kemungkinan terjadinya bias selama proses pelatihan. Hal tersebut tercermin dari nilai precision dan recall pada kedua kelas yang memiliki perbedaan sangat kecil.

Algoritma Random Forest juga memberikan pengaruh terhadap tingginya performa klasifikasi. Random Forest bekerja dengan membangun sejumlah pohon keputusan menggunakan subset data yang berbeda, kemudian menentukan hasil klasifikasi melalui mekanisme majority voting. Pendekatan ensemble tersebut mampu meningkatkan stabilitas

model, mengurangi risiko overfitting, serta menghasilkan kemampuan generalisasi yang baik pada data yang belum pernah dipelajari sebelumnya. Karakteristik tersebut menjadikan Random Forest sesuai digunakan untuk mengolah data hasil ekstraksi MFCC yang memiliki jumlah fitur relatif sedikit. Meskipun menghasilkan performa yang tinggi, model masih melakukan kesalahan klasifikasi pada sebagian kecil data pengujian. Berdasarkan hasil evaluasi, dari 900 data pengujian, sebanyak 873 data berhasil diklasifikasikan dengan benar, sedangkan 27 data mengalami kesalahan klasifikasi. Kesalahan tersebut terdiri atas 11 audio REAL yang diprediksi sebagai FAKE dan 16 audio FAKE yang diprediksi sebagai REAL. Kondisi ini menunjukkan bahwa beberapa sampel memiliki karakteristik spektral yang saling menyerupai sehingga sulit dibedakan hanya berdasarkan fitur MFCC. Selain itu, faktor seperti kualitas rekaman, keberadaan noise, variasi intonasi pembicara, serta durasi audio yang berbeda juga dapat memengaruhi proses ekstraksi fitur dan berdampak pada hasil klasifikasi.

Jika dibandingkan dengan penelitian terdahulu, pendekatan yang digunakan pada penelitian ini memiliki keunggulan dari sisi kompleksitas komputasi. Sebagian besar penelitian mengenai deteksi audio deepfake memanfaatkan model berbasis deep learning atau ensemble learning yang umumnya memerlukan sumber daya komputasi lebih besar dan waktu pelatihan yang lebih lama. Sebaliknya, penelitian ini menggunakan kombinasi MFCC dan Random Forest yang lebih sederhana, namun tetap mampu menghasilkan performa klasifikasi yang tinggi. Hal ini menunjukkan bahwa pendekatan berbasis machine learning konvensional masih relevan digunakan sebagai alternatif dalam mendeteksi audio deepfake, khususnya pada perangkat dengan kemampuan komputasi yang terbatas. Meskipun demikian, penelitian ini masih memiliki beberapa keterbatasan. Proses ekstraksi fitur hanya memanfaatkan 13 koefisien MFCC, sehingga informasi audio yang digunakan sebagai masukan model masih terbatas pada karakteristik spektral tertentu. Selain itu, penelitian ini hanya menguji satu algoritma klasifikasi, yaitu Random Forest, sehingga belum dilakukan perbandingan performa dengan algoritma machine learning lainnya maupun model deep learning. Oleh karena itu, penelitian selanjutnya dapat mengombinasikan MFCC dengan fitur audio lain seperti Chroma Feature, Spectral Centroid, Zero Crossing Rate, atau Mel Spectrogram, serta membandingkan beberapa algoritma klasifikasi untuk memperoleh performa deteksi yang lebih optimal.

Secara keseluruhan, hasil penelitian membuktikan bahwa kombinasi metode MFCC dan Random Forest mampu menjadi pendekatan yang efektif dalam mendeteksi deepfake voice. Dengan akurasi sebesar 97%, metode yang diusulkan mampu membedakan audio REAL dan audio FAKE secara akurat, sekaligus menawarkan kebutuhan komputasi yang lebih ringan dibandingkan pendekatan deep learning sehingga berpotensi diterapkan pada sistem keamanan digital, autentikasi suara, maupun forensik audio

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Penelitian ini berhasil mengimplementasikan metode Mel Frequency Cepstral Coefficients (MFCC) sebagai teknik ekstraksi fitur dan algoritma Random Forest sebagai model klasifikasi untuk mendeteksi audio REAL dan FAKE menggunakan Fake-or-Real (FoR) Dataset. Proses ekstraksi menghasilkan 13 koefisien MFCC yang mampu merepresentasikan karakteristik spektral setiap sinyal audio sehingga dapat digunakan sebagai masukan bagi model klasifikasi. Berdasarkan hasil pengujian dengan pembagian data sebesar 80% untuk pelatihan dan 20% untuk pengujian, model memperoleh akurasi sebesar 97%, disertai nilai precision, recall, dan F1-score yang tinggi dan relatif seimbang pada kedua kelas. Hasil tersebut menunjukkan bahwa kombinasi MFCC dan Random Forest mampu membedakan audio asli dan audio hasil sintesis secara efektif. Selain memberikan performa klasifikasi yang baik, metode yang digunakan juga memiliki proses pelatihan yang sederhana dan kebutuhan komputasi yang relatif rendah sehingga berpotensi diterapkan sebagai alternatif dalam sistem deteksi deepfake voice pada lingkungan dengan sumber daya komputasi yang terbatas.

5.2 Saran

Penelitian selanjutnya dapat dikembangkan dengan menambahkan fitur audio lain, seperti Chroma Feature, Spectral Centroid, Zero Crossing Rate, atau Mel Spectrogram, sehingga informasi yang digunakan dalam proses klasifikasi menjadi lebih lengkap. Selain itu, performa model dapat dibandingkan dengan algoritma machine learning maupun deep learning lainnya, seperti Support Vector Machine (SVM), XGBoost, atau Convolutional Neural Network (CNN). Pengujian menggunakan dataset yang lebih beragam, jumlah data yang lebih besar, serta kondisi audio yang mengandung noise juga perlu dilakukan agar model memiliki kemampuan generalisasi yang lebih baik dan lebih siap diterapkan pada kondisi nyata.

DAFTAR PUSTAKA

[1] Alnaqbi, M., dan Ikuesan, R.A., 2026, *A Systematic Review of Audio Deepfake Detection Techniques for Digital Investigation*, Discover Computing, Vol. 29, Article 202. DOI: **10.1007/s10791-026-10077-1**.

[2] Pham, L., Lam, P., Tran, D., Tang, H., Nguyen, T., Schindler, A., Skopik, F., Polonsky, A., dan Vu, H.C., 2025, *A Comprehensive Survey with Critical Analysis for Deepfake Speech Detection*, Computer Science Review. DOI: **10.1016/j.cosrev.2025.100757**.

[3] Yi, J., Wang, C., Tao, J., Zhang, X., Zhang, C.Y., dan Zhao, Y., 2023, *Audio Deepfake Detection: A Survey*. DOI: **10.48550/arXiv.2308.14970**.

[4] Firc, A., Malinka, K., dan Hanáček, P., 2025, *Evaluation Framework for Deepfake Speech Detection: A Comparative Study of State-of-the-Art Deepfake Speech Detectors*, Cybersecurity, Vol. 8, Article 50. DOI: 10.1186/s42400-024-00346-1.

[5] Frank, J., dan Schönherr, L., 2021, *WaveFake: A Data Set to Facilitate Audio Deepfake Detection*. DOI: **10.48550/arXiv.2111.02813**.

[6] Yamagishi, J., Wang, X., Todisco, M., Sahidullah, M., Patino, J., Nautsch, A., Liu, X., Lee, K.A., Kinnunen, T., Evans, N., dan Delgado, H., 2021, *ASVspoof 2021: Accelerating Progress in Spoofed and Deepfake Speech Detection*. DOI: 10.21437/ASVSPOOF.2021-8.

[7] Hasan, M.R., Jamil, M., Rabbani, M.G., dan Rahman, M.S., 2004, *Speaker Identification Using Mel Frequency Cepstral Coefficients*, Proceedings of the 3rd International Conference on Electrical & Computer Engineering. **DOI:** 10.1109/ICECE.2004.1399693

[8] Breiman, L., 2001, *Random Forests*, Machine Learning, Vol. 45, No. 1, pp. 5–32. DOI: **10.1023/A:1010933404324**

[9] Schonlau, M., dan Zou, R.Y., 2020, *The Random Forest Algorithm for Statistical Learning*, Stata Journal. DOI: **10.1177/1536867X20909688**.

[10] Davis, S., dan Mermelstein, P., 1980, *Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences*, IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 28, No. 4, pp. 357–366.
DOI: **10.1109/TASSP.1980.1163420**.

[11] Altalahin, I., AlZu'bi, S., Alqudah, A., dan Mughaid, A., 2023, *Unmasking the Truth: A Deep Learning Approach to Detecting Deepfake Audio Through MFCC Features*, 2023 International Conference on Information Technology (ICIT), IEEE, pp. 511–518.
DOI: **10.1109/ICIT58056.2023.10226172**.

[12] Sivaramakrishnan, M., Rajput, A., dan Saravanan, M., 2024, *Classification of Deep Fake Audio Using MFCC Technique*, 2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS).
DOI: **10.1109/ICITEICS61368.2024.10625305**.

[13] Ali, G., Rashid, J., Hussnain, M.R., Tariq, M.U., dkk., 2024, *Beyond the Illusion: Ensemble Learning for Effective Voice Deepfake Detection*, IEEE Access, Vol. 12, pp. 149940–149959.
DOI: **10.1109/ACCESS.2024.3457866**.

[14] Mcuba, M., Singh, A., Ikuesan, R.A., dan Venter, H.S., 2023, *The Effect of Deep Learning Methods on Deepfake Audio Detection for Digital Investigation*, Procedia Computer Science, Vol. 219, pp. 211–219.
DOI: **10.1016/j.procs.2023.01.283**

[15] Kamaruddin, N., dan Wahab, A., 2009, *Features Extraction for Speech Emotion*, Journal of Control and Measurement, Vol. 9, No. 1, pp. 1–12.
DOI: 10.3233/JCM-2009-0231.

[16] Hidayat, R., Bejo, A., Sumaryono, S., dan Winursito, A., 2018, *Denoising Speech for MFCC Feature Extraction Using Wavelet Transformation in Speech Recognition System*, Proceedings of the 2018 International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), IEEE.
DOI: 10.1109/ICITEED.2018.8534807.

[17] Pal, M., 2005, *Random Forest Classifier for Remote Sensing Classification*, International Journal of Remote Sensing, Vol. 26, No. 1, pp. 217–222.
DOI: 10.1080/01431160412331269698

[18] Han, K., Wang, Y., Zhang, D., Wang, G., dan Chen, X., 2022, *Deep Learning-Based Spoofed Speech Detection Using MFCC and Spectrogram Features*, IEEE Access, Vol. 10, pp. 80744–80756.
DOI: 10.1109/ACCESS.2022.3194995

[19] Probst, P., Wright, M.N., dan Boulesteix, A.L., 2019, *Hyperparameters and Tuning Strategies for Random Forest*, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, Vol. 9, No. 3.
DOI: 10.1002/widm.1301

[20] Tak, H., Patino, J., Todisco, M., Nautsch, A., Evans, N., dan Larcher, A., 2021, *End-to-End Anti-Spoofing with RawNet2*, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).
DOI: 10.1109/ICASSP39728.2021.9414234

[21] Abdeldayem, M., 2023, *The Fake-or-Real Dataset (FoR Dataset)*, Kaggle, Diakses 20 Juli 2026, https://www.kaggle.com/datasets/mohammedabdeldayem/the-fake-or-real-dataset