

Pengenalan Suara Dzikir Menggunakan Ekstraksi Fitur MFCC dan *Support Vector Machine*

Arie Rahma Nurjannah¹, Ellok Sintha Maydiana², Patmi Kasih³

^{1,2,3}Teknik Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Nusantara PGRI Kediri

E-mail: ¹arierahma3006@gmail.com, ²elloksinthamaydiana@gmail.com, ³fatkasih@gmail.com

Dzikir merupakan aktivitas ibadah yang dilakukan secara lisan dan memiliki variasi pelafalan antarindividu. Perbedaan intonasi, tempo, dan karakteristik suara tersebut dapat menyulitkan sistem komputer dalam mengenali jenis bacaan dzikir. Dzikir adalah salah satu bentuk ibadah lisan yang diucapkan berulang dengan variasi pelafalan, intonasi, serta kecepatan antarindividu. Variasi ini menimbulkan tantangan dalam proses pengenalan suara secara otomatis. Penelitian ini mengusulkan sistem pengenalan suara dzikir menggunakan ekstraksi fitur Mel Frequency Cepstral Coefficients (MFCC) dan algoritma Support Vector Machine (SVM). Pada tahap inti pengolahan sinyal, suara dzikir direkam kemudian melalui proses pra-pemrosesan berupa normalisasi amplitudo, framing, dan windowing untuk mengurangi derau. MFCC digunakan untuk mengekstraksi ciri utama dengan memetakan spektrum suara ke skala mel, menghitung energi pada filterbank mel, lalu menghasilkan koefisien cepstral yang merepresentasikan bentuk spektral suara dzikir. Selanjutnya, fitur tersebut diklasifikasikan menggunakan SVM yang bekerja dengan mencari hyperplane pemisah terbaik antar kelas dzikir melalui fungsi kernel.

Kata Kunci MFCC, pengenalan suara, Support Vector Machine, dzikir, klasifikasi suara

1. PENDAHULUAN

Pengenalan suara merupakan salah satu bidang penting dalam pengolahan sinyal digital yang bertujuan untuk memungkinkan sistem komputer mengenali dan menginterpretasikan ucapan manusia. Teknologi pengenalan suara telah banyak diterapkan pada berbagai aplikasi, seperti sistem perintah suara, autentikasi berbasis biometrik, serta interaksi manusia dan komputer [1]. Dalam konteks pengolahan sinyal suara, proses pengenalan ucapan memerlukan teknik ekstraksi ciri yang mampu merepresentasikan karakteristik suara secara efektif serta metode klasifikasi yang andal untuk membedakan pola ucapan [2]. Salah satu aktivitas yang melibatkan pengucapan berulang dengan pola lafaz tertentu adalah dzikir, yang merupakan bagian dari ibadah umat Muslim. Namun, penerapan teknologi pengenalan suara secara spesifik pada bacaan dzikir masih sangat terbatas, sehingga diperlukan penelitian untuk mengembangkan sistem pengenalan suara dzikir yang akurat dan terukur.

Dalam sistem pengenalan suara, pemilihan metode ekstraksi ciri menjadi faktor penting karena berpengaruh langsung terhadap kinerja sistem klasifikasi. Salah satu metode ekstraksi ciri yang paling banyak digunakan adalah *Mel-Frequency Cepstral Coefficients* (MFCC), karena mampu merepresentasikan karakteristik spektral sinyal suara sesuai dengan persepsi pendengaran manusia [3]. MFCC bekerja dengan memetakan frekuensi linier ke dalam skala Mel, yang meniru sensitivitas telinga manusia terhadap perubahan frekuensi rendah dibandingkan frekuensi tinggi. Berbagai penelitian menunjukkan bahwa MFCC memiliki performa yang stabil dan efektif dalam pengenalan ucapan serta pengolahan sinyal suara, baik pada lingkungan terkontrol maupun kondisi yang mengandung noise [4]. Oleh karena itu, MFCC dipilih dalam penelitian ini sebagai metode ekstraksi ciri untuk merepresentasikan pola suara pada bacaan dzikir.

Dalam praktiknya, pengenalan suara dzikir secara manual masih memiliki keterbatasan, terutama terkait konsistensi pelafalan, variasi intonasi, serta perbedaan karakteristik suara antar individu. Kondisi tersebut dapat menimbulkan kesulitan dalam membedakan jenis bacaan dzikir apabila hanya mengandalkan pendengaran manusia. Permasalahan ini mendorong perlunya pengembangan sistem pengenalan suara otomatis yang mampu mengidentifikasi bacaan dzikir secara objektif dan konsisten dengan memanfaatkan teknologi kecerdasan buatan. Beberapa penelitian terkini menunjukkan bahwa kombinasi ekstraksi ciri MFCC dan algoritma klasifikasi *Support Vector Machine* (SVM) mampu memberikan performa yang baik pada sistem pengenalan suara, khususnya pada dataset berskala kecil hingga menengah [5]. Selain itu, SVM dikenal memiliki kemampuan generalisasi yang baik serta efektif dalam menangani data berdimensi tinggi seperti fitur MFCC, sehingga cocok diterapkan pada pengenalan ucapan berbasis audio religius [6]. Oleh karena itu, penelitian ini mengusulkan penerapan MFCC dan SVM sebagai solusi untuk membangun sistem pengenalan suara dzikir yang akurat dan efisien.

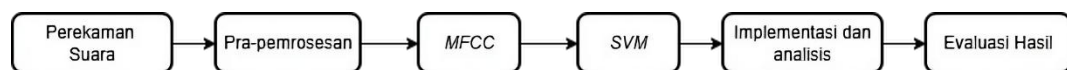
Beberapa penelitian sebelumnya juga telah mengkaji penerapan pengenalan suara pada konteks keagamaan dengan memanfaatkan metode ekstraksi ciri dan algoritma klasifikasi. Penelitian yang dilakukan oleh Sari, Fadlil, dan Sutikno [7] mengusulkan sistem pengenalan suara Islami menggunakan ekstraksi fitur *Mel-Frequency Cepstral Coefficients* (MFCC) dan algoritma *Support Vector Machine* (SVM) untuk mengklasifikasikan suara takbir dan sholawat. Hasil penelitian tersebut menunjukkan bahwa kombinasi MFCC dan SVM mampu

memberikan tingkat akurasi yang baik dalam mengenali jenis suara Islami, meskipun masih terbatas pada jumlah kelas dan variasi data suara. Selain itu, penelitian lain yang dilakukan oleh Al-Hmouz et al. [8] menunjukkan bahwa SVM memiliki performa yang stabil dalam sistem pengenalan suara berbasis MFCC, terutama pada kondisi dataset terbatas. Meskipun demikian, kajian terkait pengenalan suara dzikir secara spesifik, khususnya untuk beberapa jenis bacaan dzikir dengan karakteristik pelafalan yang mirip, masih sangat terbatas sehingga diperlukan penelitian lanjutan untuk mengembangkan sistem yang lebih terfokus dan aplikatif.

Penelitian ini mengusulkan sebuah sistem pengenalan suara dzikir menggunakan ekstraksi fitur *Mel-Frequency Cepstral Coefficients (MFCC)* dan algoritma *Support Vector Machine (SVM)*. Sistem yang dikembangkan memanfaatkan tahapan pra-pemrosesan suara yang meliputi normalisasi sinyal dan segmentasi rekaman dzikir untuk menghasilkan data audio yang lebih representatif. Setiap sinyal suara kemudian diekstraksi cirinya menggunakan MFCC dan diklasifikasikan ke dalam lima jenis bacaan dzikir, yaitu *Alhamdulillah*, *Astaghfirullah*, *Subhanallah*, *Laa Ilaaha Illallah*, dan *Allahu Akbar*. Pendekatan MFCC dan SVM dipilih karena terbukti mampu memberikan performa klasifikasi yang stabil pada pengenalan suara dengan jumlah data yang terbatas [7], [8]. Diharapkan hasil penelitian ini dapat memberikan kontribusi dalam pengembangan sistem pengenalan suara keagamaan, khususnya dzikir, serta menjadi referensi bagi penelitian selanjutnya dalam penerapan metode pengenalan suara berbasis MFCC dan SVM pada domain religius.

2. METODE PENELITIAN

Metode penelitian ini diawali dengan proses pengumpulan data berupa perekaman suara dzikir menggunakan perangkat mobile. Data suara yang diperoleh kemudian melalui tahap pra-pemrosesan untuk meningkatkan kualitas sinyal sebelum dilakukan ekstraksi ciri. Selanjutnya, ekstraksi fitur dilakukan menggunakan metode *Mel-Frequency Cepstral Coefficients (MFCC)* untuk merepresentasikan karakteristik frekuensi dari sinyal suara. Fitur yang dihasilkan digunakan sebagai masukan pada proses pelatihan dan pengujian model klasifikasi menggunakan algoritma *Support Vector Machine (SVM)*. Tahap akhir penelitian meliputi implementasi sistem serta evaluasi hasil klasifikasi menggunakan metrik evaluasi yang sesuai untuk menilai kinerja sistem secara keseluruhan. Alur umum metode penelitian ini ditunjukkan pada Gambar 2.1.



Gambar 2. 1 Metode Penelitian

Gambar 2.1 memperlihatkan alur utama metode penelitian yang terdiri atas enam tahapan, yaitu perekaman suara, pra-pemrosesan sinyal, ekstraksi ciri menggunakan MFCC, proses klasifikasi dengan SVM, implementasi sistem, serta evaluasi hasil. Pada tahap perekaman suara, data audio dzikir dikumpulkan sebagai bahan pembentukan dataset. Tahap pra-pemrosesan dilakukan untuk mereduksi derau dan menyiapkan sinyal agar layak diolah lebih lanjut. Tahap MFCC bertujuan mengekstraksi ciri dominan dari sinyal suara yang kemudian digunakan sebagai masukan pada algoritma SVM untuk proses klasifikasi jenis dzikir. Tahap implementasi menghasilkan sistem yang dapat mengenali suara dzikir secara otomatis, sedangkan tahap evaluasi digunakan untuk mengukur performa sistem yang dibangun.

2.1 Perekaman Suara

Tahap perekaman suara merupakan langkah awal dalam penelitian ini yang bertujuan untuk memperoleh data audio sebagai dasar pembentukan dataset pada proses pelatihan dan pengujian sistem klasifikasi dzikir. Perekaman dilakukan menggunakan mikrofon bawaan perangkat mobile karena mudah diakses, praktis digunakan, serta merepresentasikan kondisi nyata penggunaan aplikasi pada platform Android. Format berkas audio yang digunakan adalah MP3, mengingat format ini umum digunakan pada perangkat mobile, memiliki ukuran berkas yang relatif kecil, dan tetap mampu mempertahankan kualitas sinyal suara yang memadai untuk pengolahan sinyal digital.

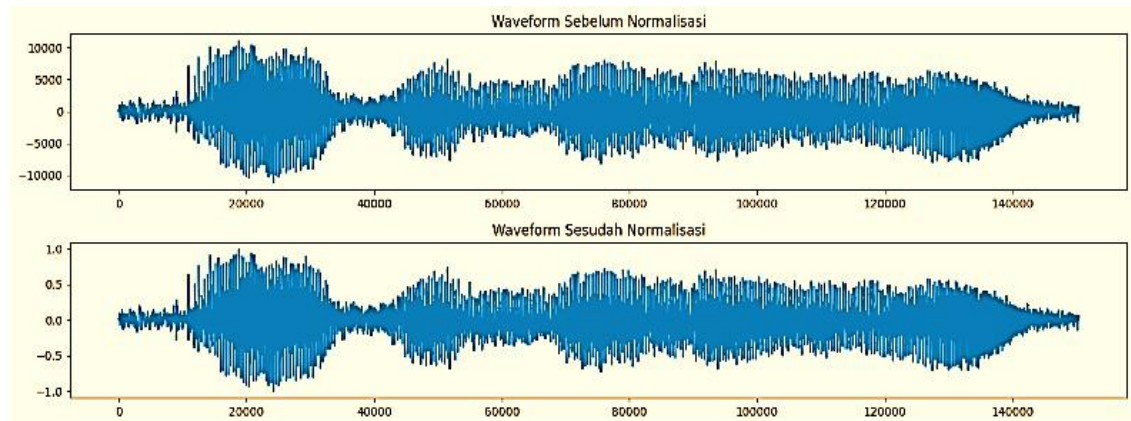
Setiap jenis dzikir direkam dalam satu berkas audio berdurasi sekitar ± 10 detik, di mana lafaz dzikir diucapkan secara berulang dalam satu sesi perekaman. Pengulangan ini bertujuan untuk memperoleh variasi data yang mencakup perbedaan pelafalan, intonasi, tempo, dan dinamika amplitudo suara, sehingga dataset yang dihasilkan lebih beragam dan representatif. Proses perekaman dilakukan di lingkungan yang relatif tenang untuk meminimalkan gangguan noise eksternal. Seluruh hasil perekaman disimpan sebagai data mentah (*raw data*) yang selanjutnya diproses pada tahap pra-pemrosesan.

2.2 Pra-pemrosesan

Tahap pra-pemrosesan bertujuan untuk menyiapkan sinyal audio agar berada pada kondisi optimal sebelum dilakukan ekstraksi ciri menggunakan *Mel-Frequency Cepstral Coefficients (MFCC)*. Kualitas sinyal hasil pra-pemrosesan sangat memengaruhi akurasi ciri yang diekstraksi serta kinerja sistem klasifikasi secara keseluruhan.

Pra-pemrosesan pada penelitian ini meliputi proses segmentasi audio, normalisasi amplitudo, dan visualisasi sinyal. Segmentasi dilakukan dengan memotong rekaman mentah berdurasi ± 10 detik menjadi beberapa segmen audio berdurasi lebih pendek, di mana setiap segmen hanya berisi satu kali pengucapan lafaz dzikir secara utuh. Melalui proses ini, satu berkas rekaman awal dapat menghasilkan beberapa berkas audio baru. Dari hasil segmentasi diperoleh sekitar 70 berkas audio untuk setiap kelas dzikir, sehingga total dataset menjadi lebih banyak dan seimbang.

Setelah segmentasi, dilakukan normalisasi amplitudo untuk menyeragamkan skala sinyal audio. Normalisasi ini bertujuan untuk mengurangi pengaruh perbedaan level suara akibat variasi jarak mikrofon dan kekuatan suara pengguna, sehingga sistem lebih fokus pada karakteristik pola frekuensi dan temporal suara. Sebagai tahap akhir, dilakukan visualisasi sinyal audio dalam bentuk waveform sebelum dan sesudah normalisasi sebagai sarana verifikasi bahwa proses pra-pemrosesan telah berjalan dengan baik dan sinyal siap digunakan pada tahap ekstraksi ciri MFCC.



Gambar 2. 2 *Waveform*

2.3 MFCC

Mel-Frequency Cepstral Coefficients (MFCC) merupakan metode ekstraksi ciri yang banyak digunakan dalam pengenalan ucapan karena mampu merepresentasikan karakteristik frekuensi yang sesuai dengan persepsi pendengaran manusia. MFCC bekerja berdasarkan konsep bahwa telinga manusia lebih sensitif terhadap perubahan frekuensi pada wilayah rendah dibanding frekuensi tinggi. Oleh karena itu, representasi frekuensi dalam MFCC menggunakan skala Mel, yaitu skala frekuensi yang didesain menyerupai cara manusia merespons suara. Konversi dari frekuensi linier ke frekuensi Mel dilakukan menggunakan persamaan:

$$mel(f) = 2595 \log_{10} \left(1 + \frac{f}{700} \right) \dots \dots \dots (2.1)$$

Tahap ekstraksi MFCC terdiri dari beberapa proses, dimulai dari *framing* untuk membagi sinyal ke dalam segmen-segmen kecil, kemudian *windowing* untuk meminimalkan diskontinuitas pada batas frame. Setiap frame selanjutnya diubah dari domain waktu ke domain frekuensi melalui *Fast Fourier Transform* (FFT). Hasil FFT kemudian dilewatkan melalui Mel Filterbank agar frekuensi yang relevan secara perseptual lebih ditekankan. Nilai energi pada setiap filter lalu dikonversi ke dalam bentuk logaritmik dan diproses menggunakan *Discrete Cosine Transform* (DCT) untuk menghasilkan koefisien MFCC. Penelitian ini menggunakan 13 koefisien MFCC yang dianggap cukup untuk mewakili karakteristik utama setiap ucapan dzikir. Hasil ekstraksi berupa matriks berukuran $13 \times n_frame$ yang menjadi representasi fitur dari setiap file audio.

2.4 Klasifikasi SVM

Support Vector Machine (SVM) merupakan metode klasifikasi yang bekerja dengan mencari *hyperplane* optimal untuk memisahkan data ke dalam beberapa kelas. *Hyperplane* tersebut direpresentasikan oleh persamaan:

$$w^T x + b = 0 \dots \dots \dots (2.2)$$

SVM menentukan *hyperplane* terbaik dengan memaksimalkan margin antara garis pemisah dan titik data terdekat. Secara matematis, optimasi SVM dirumuskan sebagai:

$$\min_{w,b} \frac{1}{2} \|w\|^2 \dots\dots\dots (2.3)$$

dengan syarat klasifikasi:

$$y_i(w^T x_i + b) \geq 1 \dots\dots\dots (2.4)$$

Untuk data yang tidak dapat dipisahkan secara linier, SVM menggunakan fungsi kernel. Pada penelitian ini digunakan kernel *Radial Basis Function* (RBF) yang didefinisikan sebagai:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \dots\dots\dots (2.5)$$

Kernel RBF memungkinkan pemetaan data ke ruang berdimensi lebih tinggi sehingga pola suara yang bersifat *non-linear* dapat dipisahkan secara optimal. Fitur yang digunakan sebagai input model berupa 13 koefisien MFCC dari setiap file audio. Selanjutnya, dataset MFCC dibagi menjadi data latih dan data uji untuk mengevaluasi performa model dalam mengenali lima jenis bacaan dzikir.

2.5 Evaluasi Kinerja Sistem

Evaluasi kinerja sistem dilakukan untuk menilai kemampuan model dalam mengklasifikasikan lima jenis bacaan dzikir. Proses evaluasi menggunakan confusion matrix, yang menyajikan informasi jumlah prediksi benar dan salah pada setiap kelas. Berdasarkan *confusion metrik* tersebut, dihitung beberapa metrik evaluasi, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*.

Accuracy digunakan untuk mengukur tingkat ketepatan prediksi model secara keseluruhan, yaitu perbandingan antara jumlah prediksi yang benar terhadap seluruh data uji. *Precision* menunjukkan tingkat ketepatan model dalam memprediksi suatu kelas tertentu, sedangkan *recall* mengukur kemampuan model dalam mengenali seluruh data yang benar-benar termasuk dalam kelas tersebut. *F1-score* merupakan nilai rata-rata harmonis antara *precision* dan *recall* yang digunakan untuk menyeimbangkan kedua metrik tersebut. *F1-score* merupakan rata-rata harmonis antara *precision* dan *recall*. Rumus masing-masing metrik ditunjukkan pada Persamaan (2.6) hingga Persamaan (2.9).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \dots\dots\dots (2.6)$$

$$Precision = \frac{TP}{TP+FP} \dots\dots\dots (2.7)$$

$$Recall = \frac{TP}{TP+FN} \dots\dots\dots (2.8)$$

$$F1 = 2 \times \frac{Precision \cdot Recall}{Precision + Recall} \dots\dots\dots (2.9)$$

3. HASIL DAN PEMBAHASAN

3.1 Hasil Pelatihan dan Pengujian Model

Pada tahap ini dilakukan pelatihan dan pengujian model klasifikasi bacaan dzikir menggunakan fitur *Mel-Frequency Cepstral Coefficients* (MFCC) dan algoritma *Support Vector Machine* (SVM). Fitur yang digunakan berupa 13 koefisien MFCC dari setiap file audio hasil pra-pemrosesan. Dataset yang telah diekstraksi kemudian dibagi menjadi data latih dan data uji dengan proporsi 80% data latih dan 20% data uji untuk mengevaluasi kemampuan model dalam mengenali lima jenis bacaan dzikir, yaitu *alhamdulillah*, *allahuakbar*, *astaghfirullah*, *lailahaillallah*, dan *subhanallah*.

Proses pelatihan model dilakukan menggunakan kernel *Radial Basis Function* (RBF) karena kernel ini mampu menangani data yang tidak terpisah secara linier, sebagaimana karakteristik umum pada sinyal suara. Model yang telah dilatih selanjutnya diuji menggunakan data uji untuk memperoleh hasil evaluasi performa secara objektif.

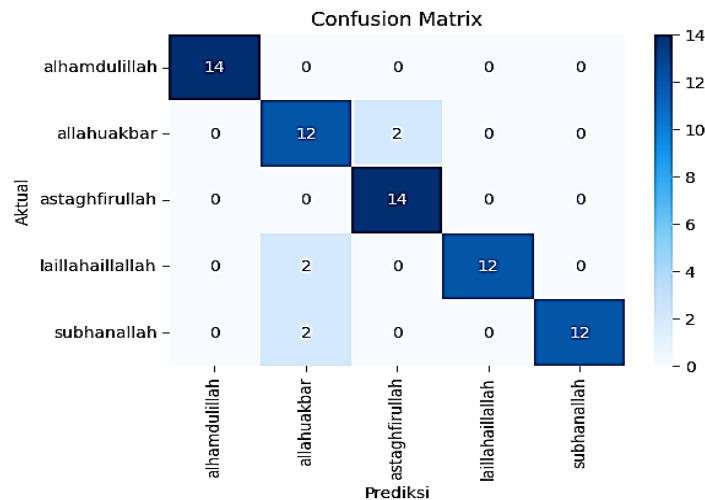
3.2 Confusion Matrix

Hasil pengujian model klasifikasi disajikan dalam bentuk *confusion matrix*, sebagaimana ditunjukkan pada Gambar 3.1. *Confusion matrix* digunakan untuk memberikan gambaran rinci mengenai performa sistem dalam mengklasifikasikan setiap kelas bacaan dzikir, baik dari sisi prediksi yang benar maupun kesalahan klasifikasi antar kelas.

Berdasarkan *confusion matrik* tersebut, dapat diketahui bahwa kelas *alhamdulillah* dan *astaghfirullah* menunjukkan performa klasifikasi yang sangat baik. Seluruh data uji pada kedua kelas tersebut berhasil diklasifikasikan dengan benar tanpa adanya kesalahan prediksi ke kelas lain. Hal ini menunjukkan bahwa karakteristik akustik dari kedua bacaan dzikir tersebut memiliki pola yang cukup khas dan dapat direpresentasikan dengan baik oleh fitur MFCC.

Pada kelas *allahuakbar*, sebagian besar data berhasil diklasifikasikan dengan benar, namun terdapat beberapa sampel yang salah diklasifikasikan sebagai kelas *astaghfirullah*. Kesalahan ini diduga disebabkan oleh kemiripan pola spektral dan dinamika suara pada beberapa bagian pengucapan, sehingga menyebabkan tumpang tindih ciri pada ruang fitur.

Kesalahan klasifikasi juga ditemukan pada kelas *lailahaillallah* dan *subhanallah*, di mana beberapa sampel diprediksi sebagai kelas *allahuakbar*. Hal ini kemungkinan disebabkan oleh kemiripan intonasi serta pola energi suara, khususnya pada bagian awal pengucapan yang memiliki karakteristik frekuensi yang serupa.



Gambar 3. 1 *Confusion Matrix*

Berdasarkan hasil pengujian yang ditunjukkan pada Gambar 3.1, kelas *alhamdulillah* dan *astaghfirullah* menunjukkan performa klasifikasi yang sangat baik.

3.3 Analisis Kinerja Sistem

Untuk memperkuat hasil evaluasi, kinerja sistem juga dianalisis secara kuantitatif menggunakan beberapa metrik evaluasi, yaitu akurasi, *precision*, *recall*, dan *F1-score*. Ringkasan hasil evaluasi kuantitatif disajikan pada Tabel 3.1.

Tabel 3. 1 Tabel Metrik Evaluasi

Metrik	Nilai
Akurasi	91,43%
<i>Precision</i>	92,50%
<i>Recall</i>	91,43%
<i>F1-score</i>	91,59%

Hasil evaluasi pada Tabel 3.1 menunjukkan bahwa sistem klasifikasi bacaan dzikir memiliki tingkat performa yang tinggi dan stabil. Nilai *F1-score* yang mendekati nilai akurasi menunjukkan keseimbangan yang baik antara *precision* dan *recall*, sehingga sistem tidak hanya akurat dalam memprediksi kelas, tetapi juga andal dalam mengenali seluruh jenis bacaan dzikir yang diuji.

4. SIMPULAN

Berdasarkan serangkaian proses yang dilakukan dalam penelitian ini, mulai dari analisis algoritma, perekaman data, implementasi sistem hingga analisis dan evaluasi hasil uji, diperoleh kesimpulan sebagai berikut:

1. Penelitian telah berhasil mengembangkan sebuah sistem klasifikasi bacaan dzikir yang mampu mengenali dan membedakan lima kelas bacaan, yaitu *alhamdulillah*, *allahuakbar*, *astaghfirullah*, *lailahaillallah*, dan *subhanallah*. Sistem yang dibangun memanfaatkan metode *Mel-Frequency Cepstral Coefficients* (MFCC) sebagai teknik ekstraksi ciri suara serta *Support Vector Machine* (SVM) sebagai algoritma klasifikasi. Kombinasi kedua metode tersebut dipilih karena MFCC efektif dalam merepresentasikan karakteristik sinyal suara sesuai dengan persepsi pendengaran manusia, sementara SVM memiliki kemampuan yang baik dalam memisahkan data antar kelas pada ruang fitur berdimensi tinggi.
2. Tahapan pra-pemrosesan yang diterapkan, meliputi segmentasi audio dan normalisasi amplitudo, terbukti berperan penting dalam meningkatkan kualitas data suara. Normalisasi amplitudo mampu mengurangi perbedaan tingkat volume akibat variasi pengucapan dan kondisi perekaman, sehingga sinyal audio menjadi lebih konsisten dan stabil. Dengan sinyal yang telah distandarkan, proses ekstraksi ciri MFCC dapat menghasilkan fitur yang lebih representatif dan mendukung kinerja klasifikasi secara optimal.
3. Hasil ekstraksi ciri menggunakan 13 koefisien MFCC menunjukkan bahwa fitur yang dihasilkan mampu menangkap karakteristik frekuensi dan pola akustik dari masing-masing bacaan dzikir dengan baik. Fitur MFCC tersebut digunakan sebagai masukan pada proses pelatihan model SVM, sehingga sistem mampu mempelajari perbedaan pola suara antar kelas secara efektif.
4. Berdasarkan hasil pengujian, model SVM menunjukkan kinerja klasifikasi yang baik dengan tingkat akurasi yang tinggi dan tingkat kesalahan klasifikasi yang relatif kecil. Evaluasi menggunakan *confusion matrix* memperlihatkan bahwa sebagian besar data uji berhasil diklasifikasikan ke dalam kelas yang sesuai. Hasil ini menunjukkan bahwa sistem yang dikembangkan memiliki tingkat keandalan yang memadai dan berpotensi untuk diterapkan sebagai sistem pendukung pengenalan bacaan dzikir berbasis suara. Meskipun demikian, penelitian ini masih memiliki keterbatasan, khususnya pada bacaan dzikir yang memiliki kemiripan karakteristik akustik dan fonetik, sehingga berpotensi menimbulkan kesalahan klasifikasi. Oleh karena itu, diperlukan pengembangan lebih lanjut untuk meningkatkan akurasi dan ketahanan sistem terhadap variasi pengucapan di kondisi nyata.

5. SARAN

Untuk penelitian selanjutnya, disarankan agar jumlah dan variasi dataset ditingkatkan dengan melibatkan lebih banyak pembicara dari berbagai latar belakang, seperti perbedaan usia, jenis kelamin, serta karakteristik suara. Selain itu, variasi kondisi perekaman, termasuk lingkungan yang lebih beragam dan penggunaan perangkat perekam yang berbeda, perlu dipertimbangkan agar sistem memiliki kemampuan generalisasi yang lebih baik dan lebih robust terhadap gangguan noise.

Pengembangan selanjutnya juga dapat difokuskan pada peningkatan metode ekstraksi ciri, seperti penambahan fitur turunan MFCC (delta dan delta-delta MFCC) maupun fitur spektral lain untuk memperkaya representasi sinyal suara. Di sisi klasifikasi, penelitian mendatang dapat melakukan perbandingan dengan algoritma lain, seperti *K-Nearest Neighbor*, *Random Forest*, atau pendekatan berbasis *deep learning* untuk memperoleh performa yang lebih optimal.

Selain itu, sistem yang dikembangkan memiliki potensi untuk diimplementasikan dalam bentuk aplikasi mobile berbasis *real-time*, sehingga pengguna dapat langsung memperoleh umpan balik terhadap bacaan dzikir yang diucapkan. Pengembangan tersebut diharapkan dapat meningkatkan manfaat sistem sebagai media bantu ibadah dan pembelajaran pelafalan dzikir yang benar.

DAFTAR PUSTAKA

- [1] L. R. Rabiner and B. H. Juang, *Fundamentals of Speech Recognition*. in Prentice-Hall Signal Processing Series: Advanced monographs. PTR Prentice Hall, 1993. [Online]. Available: <https://books.google.co.id/books?id=XEVqQgAACAAJ>
- [2] S. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE Trans Acoust*, vol. 28, no. 4, pp. 357–366, 1980, doi: 10.1109/TASSP.1980.1163420.
- [3] T. F. Quatieri, *Discrete-time Speech Signal Processing: Principles and Practice*. in Pearson Education signal processing series. Prentice Hall PTR, 2002. [Online]. Available: <https://books.google.co.id/books?id=5KYeAQAAIAAJ>
- [4] X. Huang, A. Acero, and H.-W. Hon, *Spoken language processing : a guide to theory, algorithm, and system development*. New Jersey, United States: Upper Saddle River, NJ : Prentice Hall PTR, 2001. [Online]. Available: <https://searchworks.stanford.edu/view/4683582>

- [5] R. Rosdiana, M. Ula, and H. A. K. Aidilof, “Implementasi Pemodelan Citra Model Svm (Support Vector Machine) Dalam Penentuan Pengklasifikasian Jenis Suara Kontes Burung,” *Jurnal Informatika Kaputama (JIK)*, vol. 5, no. 2, pp. 317–324, 2021.
- [6] Md. Sahidullah and G. Saha, “Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition,” *Speech Commun*, vol. 54, no. 4, pp. 543–565, 2012, doi: <https://doi.org/10.1016/j.specom.2011.11.004>.
- [7] I. P. S. Sari, A. Fadlil, and T. Sutikno, “Islamic Sound Recognition Using MFCC and SVM: Case Study on Takbir and Sholawat,” *Journal of Computer Networks, Architecture and High Performance Computing*, vol. 7, no. 3, pp. 825–839, 2025.
- [8] J. B and R. Babu.N, *Speech recognition using MFCC and DTW*. 2014. doi: 10.1109/ICAEE.2014.6838564.