Jurnal Teknologi Informasi, Komputer dan Aplikasinya (JTIKA)
Vol. 4, No. 2, September 2022, (Terakreditasi Sinta-4, SK No:164/E/KPT/2021) ISSN:2657-0327

VERIFIKASI SUARA MAHASISWA SEBAGAI ALTERNATIF PRESENSI KEHADIRAN MENGGUNAKAN EKSTRAKSI FITUR MFCC DAN KLASIFIKASI LVQ

(*Student Voice Verification As Alternative Attendance Presence Using MFCC Feature Extraction And LVQ Classification***)**

Muhammad Afif Ma'ruf, Arik Aranta*, Fitri Bimantoro
Dept Informatics Engineering, Mataram University
Jl. Majapahit 62, Mataram, Lombok NTB, INDONESIA
Email: maruf.muhammadafif@gmail.com, [arikaranta, bimo]@unram.ac.id

Abstract

Attendance is an essential thing in the learning process. In recent years, technology development has been relatively rapid, one of which is in attendance recording. Attendance registration can now be done using QR-Code, palm, face recognition and digital signature. There are shortcomings, such as the lack of flexibility in the attendance process and the problem of the pandemic being limited by distance. The existence of the teacher's online method makes it challenging to communicate. Based on these problems, this study was presented to build a voice verification model using the Mel-Frequency Cepstral Coefficients (MFCC) and Learning Vector Quantization (LVQ) methods as an alternative attendance. This study uses text-dependent recorded data from 35 speakers. In this study, verification was carried out with the condition of using a mask and without a mask. This study obtained an average margin of similarity score of 80% of the average similarity score of 93.96% native speakers and 67.58% fake speakers. The best test results were obtained at a threshold of 0.85 with an accuracy value of 86.2%, precision of 87.86%, and recall of 84% when using a mask, while without a mask the value of accuracy was 90%, precision was 88.97%, and recall was 91.17%.

Keywords: *Presensi, Text Dependent, Verifikasi Pembicara, MFCC, LVQ*

**Penulis Korespondensi*

1. PENDAHULUAN

Dalam beberapa tahun terakhir, *speech recognition* mempunyai peranan penting dalam perkembangan teknologi. Beberapa perusahaan teknologi seperti Google, Samsung, dan Apple telah menerapkan *speech recognition* untuk menerjemahkan ucapan manusia menjadi perintah untuk membantu menggunakan produk dengan mudah [1]. Pengenalan pembicara atau *speaker recognition* adalah sebuah metode yang digunakan untuk mengenali identitas seseorang secara otomatis dengan menggunakan komputer.

Pesatnya perkembangan teknologi mendukung adanya perkembangan dalam proses presensi. Pencatatan kehadiran mahasiswa dapat dilakukan menggunakan QR-Code, telapak tangan, pengenalan wajah dan tanda tangan digital [2][3]. Namun dalam pelaksanaannya ada kekurangan seperti kurangnya fleksibilitas dalam proses presensi[4]. Permasalahan juga timbul di masa pandemi dengan adanya metode pengajaran secara jarak jauh, guru hampir tidak melihat mahasiswa dan lebih sulit untuk mengingat wajah atau suara mahasiswa. Beberapa mahasiswa juga tidak memiliki akses internet dan kamera yang memadai baik selama pelajaran, latihan dan ujian [5].

Berjalannya waktu pembelajaran kembali secara bertahap menggunakan pembelajaran secara luring dengan aturan penggunaan masker secara ketat. Faktor COVID-19 membuat sebagian masyarakat termasuk dunia pendidikan wajib menggunakan masker di saat berkegiatan. Penggunaan masker menghadirkan faktor eksternal yang merugikan proses presensi seperti komunikasi ucapan [6].

Verifikasi pembicara atau *speaker verification* dapat digunakan sebagai alternatif untuk memvalidasi presensi kehadiran. Verifikasi dilakukan untuk memvalidasi identitas pembicara sehingga dapat diberikan kesimpulan diterima atau ditolaknya sebagai pembicara yang sebenarnya. Informasi pembicara berupa jenis kelamin, usia, dan identitas individu dapat ditemukan pada karakteristik manusia yang disebut biometrik [7].

Biometrik juga bisa didapat pada suara karena sinyal ucapan manusia merupakan sinyal multidimensi yang membawa berbagai informasi seperti frekuensi dasar (*Pitch*), energi dan durasi pengucapan sehingga memiliki keunikan satu individu dengan individu lain. Biometrik Suara dapat dijadikan alternatif verifikasi identitas, meningkatkan fleksibilitas dan dapat menambah kekayaan ilmu pengetahuan terkait penerapan kecerdasan buatan dalam proses presensi kehadiran.

Dalam memperoleh informasi biometrik dari suara manusia, *Mel Frequency Cepstral Coefficients* (MFCC) menjadi metode yang sering digunakan dalam bidang ekstraksi fitur suara. MFCC merupakan metode yang menghitung koefisien *cepstral* berdasarkan variasi frekuensi kritis pada sistem pendengaran manusia, sehingga mampu merepresentasikan sinyal suara sebagaimana manusia mendengar [8]. Penelitian sebelumnya mengenai penggunaan ekstraksi fitur MFCC untuk identifikasi pembicara didapatkan akurasi *identification rate* sebesar 88.9% [8]. Berdasarkan penjelasan tersebut MFCC dapat digunakan sebagai ekstraksi fitur biometrik suara.

Pendekatan yang dapat digunakan untuk mengklasifikasikan informasi sinyal suara salah satunya menggunakan *neural network*. LVQ merupakan salah satu dari jenis *neural network* yang menggunakan pembelajaran *supervised learning*. Pembelajar LVQ cukup ideal karena masukan yang memiliki kemiripan akan dikelompokkan, sehingga hasil yang diberikan pada proses pembelajaran akan memiliki hasil yang lebih akurat. Tujuan penggunaan LVQ untuk mengelompokkan masukan terhadap keluaran dalam klasifikasi vektor agar dapat meminimalkan proses terjadinya kesalahan dalam klasifikasi [9].

Berdasarkan pemaparan di atas, penulis mengajukan untuk melakukan penelitian membuat model verifikasi presensi kehadiran mahasiswa berbasis suara menggunakan metode MFCC dan LVQ. Penelitian ini bertujuan untuk menguji penerapan metode MFCC dan LVQ dalam verifikasi berbasis suara.

2. TINJAUAN PUSTAKA

Pada penelitian sejenis, yaitu pengenalan suara untuk identifikasi personal menggunakan ekstraksi fitur MFCC mendapatkan akurasi sebesar 82,67%. Penelitian tersebut menggunakan data rekaman suara pengucapan nama dengan durasi kurang dari tiga detik, dengan total pembicara lima orang [10].

Telah dilakukan penelitian dengan pengenalan suara dengan studi kasus dosen menggunakan MFCC dan CNN. Data menggunakan rekaman *voice note* dari 4 dosen, masing-masing memiliki durasi 5 detik dengan frekuensi 44kHz, serta menggunakan *cross validation* mendapatkan akurasi sebesar 93,75% dan 100% [11].

Penelitian dengan menggunakan bahasa arab untuk identifikasi pembicara mendapatkan akurasi sebesar 97.5%. Penelitian tersebut menggunakan *pre-emphasis* 0.97 dengan jumlah MFCC sebanyak 28 [7].

Penelitian menggunakan MFCC telah dilakukan dengan menggabungkan dengan GMM untuk mengenali pembicara, dan didapatkan akurasi sebesar 94.12%. Penelitian tersebut menggunakan 15 pembicara sebagai data dan dilakukan dua pengujian yaitu pada *text independent* dan *text dependent* [12].

Penelitian di luar dari identifikasi pembicara telah dilakukan yaitu mengenali penyebutan bahasa Inggris. Penelitian tersebut mencoba mengenali 4 jenis kata menggunakan *text to speech converter* dan mendapatkan akurasi sebesar 80%, presisi 85% dan recall 92% [13].

Penggunaan klasifikasi LVQ digunakan pada *text dependent* untuk identifikasi suara multilingual dengan data set terdiri dari 40 pembicara dengan 3 bahasa yaitu *english*, *sanskrit*, dan *hindi*. Menggunakan 20ms *frame*, 50% *overlap*, 35 *coefficients* dan mendapatkan akurasi sebesar 80.52% [14].

Berdasarkan penelitian-penelitian tersebut, dapat diketahui bahwa metode ekstraksi fitur MFCC memiliki nilai ekstraksi fitur yang cukup baik untuk diterapkan dalam pengenalan kata dan pengenalan pembicara. Metode klasifikasi LVQ juga memiliki hasil tingkat akurasi yang baik untuk pengklasifikasian data sinyal suara. Oleh karena itu, penulis mengajukan penelitian verifikasi presensi suara menggunakan ekstraksi fitur MFCC dan klasifikasi LVQ.

2.1. Sinyal Suara

Sinyal suara merupakan sinyal quasi-stationary, yaitu ketika diperiksa selama periode 5-100 milidetik karakteristiknya cukup stasioner, namun dalam jangak waktu yang lebih panjang karakteristik sinyal berubah mencerminkan masukan yang diucapkan [15]. Dalam persepsi pendengaran manusia, sinyal suara memiliki nada dengan frekuensi dan pola tertentu yang dapat diukur dengan skala mel. Skala mel merupakan rentan di bawah 1000Hz dan skala logaritmik di atas 1000Hz [16].

2.2. Biometrik Suara

Otentikasi suara didasari bahwa setiap manusia berbeda dalam nada, volume dan volume yang membuatnya dapat dibedakan secara unik. Faktor yang memengaruhi suara berbeda beda satu dengan yang lain adalah ukuran dan bentuk mulut,

tenggorokan, hidung dan gigi serta ukuran, bentuk dan tebal pita suara [17].

Sinyal ucapan manusia merupakan sinyal multidimensi yang membawa berbagai informasi seperti frekuensi dasar (*Pitch*), energi dan durasi pengucapan sehingga memiliki keunikan satu individu dengan individu lain dan peluang bahwa semua hal tersebut persis sama pada dua orang sangat rendah [18]. Biometrik suara memiliki keunggulan seperti sinyal alami yang diproduksi tubuh, biaya implementasi yang rendah dan memungkinkan pengguna untuk melakukan otentikasi jarak jauh.

2.3. File WAV

File WAV (*Waveform Audio File Format*) merupakan format *file digital audio* yang disimpan dengan ekstensi WAV. File tersebut menyimpan amplitudo dan frekuensi sinyal suara tidak terkompresi sehingga *file* yang dihasilkan memiliki penyimpanan besar. File WAV memiliki informasi berupa ekstensi *file* WAV, *samplerate* dan berisi data sinyal tanpa kompresi [19].

2.4. Presensi

Presensi adalah sebuah kegiatan pengambilan data yang digunakan untuk mencatat jumlah kehadiran mahasiswa dalam kegiatan. Kemudian diproses untuk mendapatkan informasi sebagai tolak ukur kesuksesan kegiatan atau sebuah informasi dari yang telah mengikuti kegiatan tersebut. Presensi dalam penelitian ini ditujukan untuk mencari tahu kehadiran mahasiswa yang datang pada kegiatan. Pengambilan data presensi secara manual memiliki banyak kekurangan seperti adanya data salah masuk, data hilang, data rusak dan faktor pengambilan presensi yang lama sehingga mengurangi efisiensi dan efektivitas presensi [4].

2.5. Identifikasi Pembicara

Identifikasi suara adalah proses menentukan pembicara yang berbicara. Dalam proses identifikasi pembicara, jumlah keputusan tergantung pada jumlah orang dalam database yang digunakan, sehingga kinerja sistem identifikasi suara akan menurun, jika ukuran suara yang digunakan dalam sistem meningkat. Secara umum pernyataan ucapan akan diproses dan dianalisis untuk dibandingkan dengan berbagai model pembicara, kemudian suara yang diberikan akan diidentifikasi sebagai pembicara sesuai dengan model yang diidentifikasi [20].

2.6. Autentikasi Suara

Autentikasi suara adalah proses verifikasi identitas pembicara berdasarkan yang diucapkan. Ucapan dari pembicara yang tidak dikenal dibandingkan dengan model pembicara, jika melampaui ambang batas, identitas yang diklaim verifikasi dan diterima. Jika tidak maka identitas pembicara ditolak. Penentuan ambang batas optimal untuk menerima dan menolak pembicara merupakan hal penting yang harus diperiksa oleh pembicara. Memilih ambang batas tinggi akan menghasilkan sebagian besar pembicara yang tidak dikenal ditolak, tetapi juga meningkatkan risiko menolak pembicara yang dikenal dalam sistem. Berlaku sebaliknya, oleh karena itu ambang batas optimal perlu diperhitungkan berdasarkan distribusi autentikasi pembicara yang dikenal dan tidak dikenal dalam sistem [20]

2.7. Verifikasi Pembicara

Verifikasi pembicara adalah proses menerima atau menolak klaim identitas pembicara. Tujuannya untuk mendapatkan keluaran menerima atau menolak suatu identitas yang diklaim oleh pembicara. Skor verifikasi dapati menggunakan model pembicara yang diklaim dan model anti-pembicara, skor tersebut dibandingkan dengan ambang batas. Jika skor lebih tinggi dari ambang batas, pembicara diterima dan sebaliknya. Skema verifikasi pembicara dapat digambarkan sebagai berikut [21].

Gambar 1. Skema sistem verifikasi pembicara

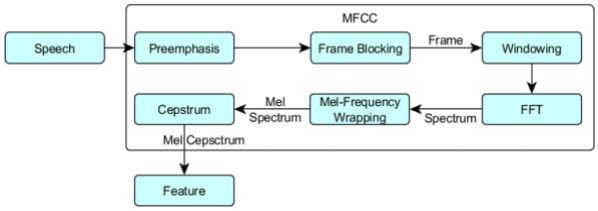
2.8. *Text Dependent*

Verifikasi pembicara Proses pengenalan pembicara dengan masukan suara dapat dibagi menjadi dua bagian. Salah satunya pengenalan pembicara dengan menggunakan teks suara yang sama saat *training phase* dan *testing phase* disebut dengan *text-dependent speaker*, tipe ini tidak kompleks karena menggunakan beberapa kata sebagai sampel [8][15]. Metode tersebut biasanya menggunakan teknik pencocokan *template* yang dilihat dari sumbu waktu dan ucapan sampel masukan terhadap referensi pembicara yang kemudian dihitung kesamaan dua suara [22].

2.9. *Mel Frequency Cepstral Coefficients* (MFCC)

Mel Frequency Cepstral Coefficients (MFCC) banyak digunakan pada penelitian yang berhubungan dengan suara manusia. Ekstraksi ciri sinyal suara menggunakan MFCC didasarkan atas variasi *bandwith* terhadap frekuensi pada telinga manusia yang bekerja secara linear pada frekuensi rendah dan bekerja secara logaritmik pada frekuensi tinggi. Filter ini dapat

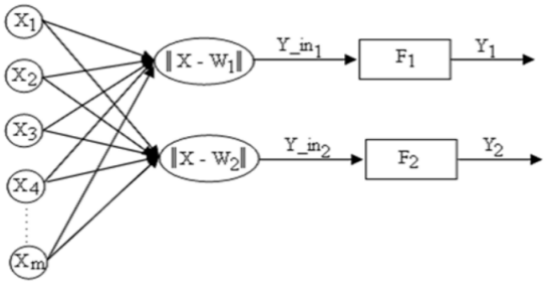
digunakan untuk menangkap informasi penting dari sinyal suara masukan atau ucapan [15]. Karakteristik filter digambarkan dalam sekala *mel*-frekuensi dengan frekuensi linear di bawah 1000Hz dan frekuensi logaritmik di atas 1000Hz. Proses MFCC dapat dilihat pada gambar berikut.



Gambar 2. Blok diagram proses MFCC

2.10. *Learning Vector Quantization* (LVQ)

Learning Vector Quantization (LVQ) merupakan jaringan syaraf tiruan dengan tipe arsitektur jaringan lapis tunggal *Single Layer Feedforward*. LVQ bertujuan untuk mendapatkan pembelajaran pada lapisan kompetitif yang terawasi. Lapisan kompetitif ini secara otomatis dapat mempelajari masukan untuk mengklasifikasikan vektor berdasarkan jarak antara vektor masukan dan bobot menggunakan *Euclidian distance*. Arsitektur LVQ dapat dilihat pada Gambar 3 [13].



Gambar 3. Arsitektur *Learning Vector Quantization*.

Algoritma dalam melakukan LVQ adalah sebagai berikut [23]:

1. Tentukan maksimum *epoch* (banyak langkah proses pelatihan yang berulang) dan nilai alpha.
2. Hasil ekstraksi fitur pertama dari masing-masing pola digunakan sebagai data awal. Data awal akan diisi sebagai nilai bobot awal (*w*).
3. Nilai *Epoch* dimulai dari 0.
4. Selama (*Epoch* < *MaxEpoch*), maka lanjut ke langkah selanjutnya
5. *Epoch* = *Epoch* + 1
6. Untuk setiap data hasil ekstraksi fitur, lakukan hal berikut:
   a. Set x = hasil ekstraksi fitur dari pola.
   b. Set T = nomor urut dari setiap kelas.
   c. Hitung jarak hasil ekstraksi fitur pola saat ini dengan masing-masing bobot. Misalkan dihitung jarak hasil ekstraksi fitur pertama dengan setiap bobot menggunakan persamaan (1)

$$J = \sqrt{(x_{11} - w_{11})^2 + \cdots + (x_{1m} - w_{1m})^2} \quad (1)$$

   d. Bila nomor kelas pada bobot yang memiliki jarak terkecil sama dengan nilai nomor urut (T) pol, maka hitung:

$$w_j(baru) = w_j(lama) + \alpha(x - w_j(lama)) \quad (2)$$

   e. Bila tidak, maka hitung:

$$w_j(baru) = w_j(lama) - \alpha(x - w_j(lama)) \quad (3)$$

7. Kurangi nilai α

$$\alpha = \alpha - (0.1 * \alpha) \quad (4)$$

2.11. *Cosine Distance Feature* (CDF)

Metode *cosine distance feature* merupakan metode penilaian yang populer dan efisien secara komputasi untuk verifikasi pembicara. Metode ini menggunakan pendekatan *cosine distance* antara vektor-i pembicara data latih dan vektor pembicara data uji. Nilai tersebut yang digunakan sebagai skor keputusan. Perumusan nila *cosine distance feature* ditunjukkan pada persamaan 5 [24][21].

$$CDF(vec_{tar}, vec_{test}) = \frac{vec_{tar} . vec_{test}}{|vec_{tar}||vec_{test}|} \quad (5)$$

Cosine distance dalam penelitian sebelumnya mendapatkan nilai *Identification Rate* terbaik dari *distance* lainnya dalam pengenalan suara [21].

2.12. Evaluasi Hasil

Evaluasi hasil dapat dilakukan menggunakan metrik evaluasi. Metrik evaluasi dihitung b nilai *True Positive, True Negative, False Positive dan False Negative*. Evaluasi tersebut dapat dihitung berdasarkan *confussion matriks* seperti pada Tabel I [25]. Hasil dari *confussion matrix* kemudian akan digunakan untuk menghitung evaluasi hasil klasifikasi

TABEL I. *CONFUSSION MATRIX*

Hasil Klasifikasi / Kelas Sebenarnya	**Positif**	**Negatif**
Positif	TAPI	FP
Negatif	FN	TN

Dari *confussion matrix* dapat dilakukan perhitungan evaluasi berupa nilai akurasi, presisi dan *recall*. Akurasi merupakan rasio prediksi benar (positif dan negatif) dengan jumlah keseluruhan total data.

Perhitungan akurasi dapat menggunakan persamaan berikut.

$$akurasi = \frac{TP+TN}{TP+FN+FP+TN} \quad (6)$$

Presisi merupakan rasio prediksi benar positif dibandingkan dengan jumlah keseluruhan hasil yang diprediksi positif. Perhitungan presisi dapat menggunakan persamaan berikut.

$$presisi = \frac{TP}{TP+FP} \quad (7)$$

Recall merupakan rasio prediksi benar positif dibandingkan dengan jumlah keseluruhan hasil data yang benar positif. Perhitungan *recall* menggunakan persamaan berikut.

$$recall = \frac{TP}{TP+FN} \quad (8)$$

3. METODE PENELITIAN

3.1. Alat dan Bahan

Alat - alat yang diperlukan dalam penelitian ini dibagi menjadi dua yakni perangkat keras dan perangkat lunak antara lain sebagai berikut:

1. Perangkat keras

   Perangkat keras yang digunakan dalam penelitian ini adalah perangkat dengan spesifikasi sebagai berikut:

   a. Laptop Asus UX430UNR Intel® Core™ i5-8520U dengan GPU NVIDIA GEFORCE MX150, dan RAM 8 GB.
   b. *Smartphone* masing-masing pembicara.
2. Perangkat Lunak
   a. Sistem operasi Windows
   b. Jupyter Lab
   c. Bahasa pemrograman Python versi 3.

3.2. Data Penelitian

Bahan penelitian yang digunakan dalam penelitian ini adalah rekaman suara dalam format ".wav" sebanyak 1375 data. Komposisi data yaitu 25 orang pembicara sebagai pembicara asli dan 10 orang pembicara sebagai pembuat rekaman menirukan pembicara asli. Data suara diambil suara laki – laki dan perempuan. Data suara diambil dengan durasi rata-rata 2 sampai 3 detik, pembagian data terdiri dari data latih, data klaim (tanpa masker dan masker), dan data uji. Menggunakan *device* masing-masing pembicara [26].

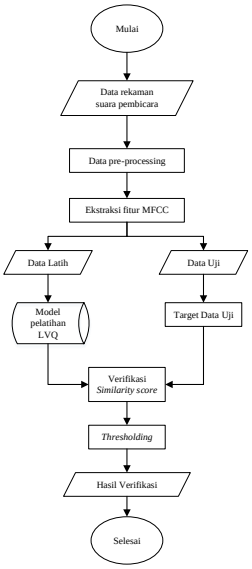
Jumlah pembicara yang digunakan dalam penelitian ini berdasarkan beberapa penelitian sebelumnya sebagai acuan. Pada penelitian identifikasi pembicara menggunakan 20 pembicara [27]. Penelitian sebelumnya dengan studi kasus yang sama yaitu kehadiran siswa, menggunakan 26 pembicara berbeda sebagai data latih dan data uji [28].

Gambar 4. Data yang sudah dipilah dalam folder.

Setelah data suara dimasukkan ke dalam folder, selanjutnya data suara dimasukkan dalam tahap segmentasi suara untuk menghilangkan *silent* jeda di awal dan di akhir rekaman.

3.3. Perancangan Sistem

Dalam pembangunan model pembelajaran menggunakan fitur MFCC dengan klasifikasi LVQ, terdapat beberapa tahapan yang akan dilakukan, dimulai dari segmentasi data, ekstraksi ciri, pembagian data, dan pembuatan model LVQ. Proses tersebut dijabarkan dalam diagram alir pembangunan model pada gambar 5.



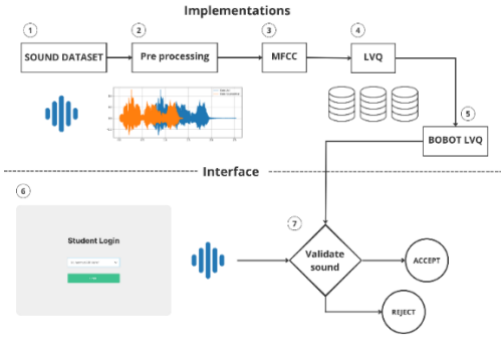
Gambar 5. Diagram alir perancangan model.

Seperti terlihat pada gambar 5, rekaman data akan dilakukan segmentasi suara dengan menghilangkan suara *silent* jeda di awal dan di akhir. Kemudian akan dilanjutkan dengan ekstraksi fitur MFCC untuk seluruh data yang hasilnya akan dilakukan proses perhitungan bobot dalam klasifikasi LVQ. Bobot model yang sudah dihitung akan dilakukan verifikasi menggunakan *similarity score* menggunakan perhitungan CDF. Perhitungan CDF dilakukan dengan mengelompokkan target tertentu sehingga verifikasi akan dilakukan satu target data yang diujikan dengan bobot target yang

sama. Terakhir yaitu proses penentuan *threshold* dengan melihat hasil terbaik dari verifikasi yang dilakukan.

3.4. Implementasi Sistem

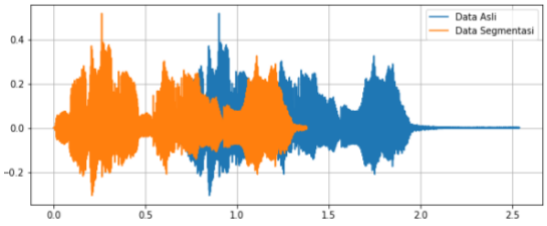
Pada penelitian ini dibuatkan perencanaan implementasi sistem sebagai informasi penerapan model di dalam melakukan verifikasi absensi mahasiswa menggunakan media suara. seperti pada gambar 6 dapat dilihat bahwa proses dimulai melakukan *pre-processing* pada seluruh dataset, kemudian dilakukan ekstraksi fitur dan pembelajaran menggunakan *learning vector quantization*. Bobot terlatih ini yang kemudian akan dibandingkan dengan masukan suara pembicara. Pertama memilih pembicara yang akan di akses, kemudian suara dimasukkan dan dibandingkan dengan bobot sesuai dengan pembicara yang dipilih. Hasil keluarannya adalah diterima atau ditolak.



Gambar 6. Ilustrasi implementasi sistem.

3.5. *Pre-processing*

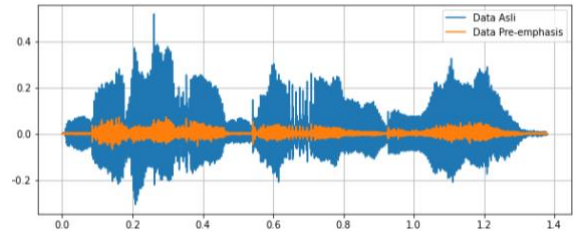
Pada penelitian ini proses *pre-processing* yang dilakukan adalah segmentasi atau pemotongan suara pada rekaman dengan kondisi suara *silent* di awal dan akhir rekaman suara pada semua data. Proses segmentasi dilakukan menggunakan *library* Pydub *audio*, segmentasi dilakukan untuk mendapatkan rekaman suara tanpa adanya jeda awal dan akhir ketika proses perekaman suara. Contoh sampel dari *pre-processing* disajikan pada Gambar 6.



Gambar 7. Hasil proses *pre-processing*.

3.6. *Pre-emphasis*

Pada *pre-emphasis* hasil dari *pre-processing* dilakukan untuk meningkatkan energi dari frekuensi tinggi agar sama dengan energi frekuensi rendah. Perbedaan ini terjadi karena frekuensi rendah di *sampling* dengan frekuensi yang cukup tinggi untuk mendapatkan informasi dari suara. Nilai parameter *alpha* pada *pre-emphasis* bernilai 0.97. Nilai tersebut sering digunakan para proses *pre-emphasis* sinyal suara [29]. Parameter tersebut digunakan pada semua *dataset* sebelum dilakukan ekstraksi fitur seperti pada Gambar 7.



Gambar 8. Hasil proses *pre-emphasis*.

3.7. Ekstraksi fitur

Pada tahapan ini digunakan ekstraksi fitur MFCC dengan menggunakan *library* Librosa pada Python dengan fungsi librosa.feature.mfcc. Pada fungsi tersebut terdapat beberapa parameter yang digunakan dalam skenario pengujian, yaitu panjang *frame*, panjang *overlap* dan jumlah koefisien MFCC. Tahap ekstraksi fitur dilakukan kepada seluruh data kemudian hasil fitur-fitur tersebut digunakan dalam proses LVQ dan verifikasi pembicara.

3.8. Skenario pengujian

Pengujian skenario dilakukan dengan tujuan untuk menguji model yang dihasilkan dengan memperhatikan parameter – parameter uji coba. Pada penelitian ini akan dilakukan beberapa tahapan skenario pengujian seperti berikut.

1. Pengujian parameter LVQ dengan variasi jumlah *epoch* 50, 100, 200 serta pengaruh *learning rate* dengan rentang 0.1, 0.01, dan 0.001.
2. Pengujian parameter *threshold* dengan rentang 0.5 sampai 0.9.
3. Pengujian panjang *frame* setiap rekaman suara sebesar 20ms, 25ms, dan 30ms dengan jumlah koefisien ekstraksi fitur MFCC sebesar 12, 20, dan 40.
4. Pengujian pengaruh *similarity score* data pembicara asli dan peniru suara pembicara asli.
5. Pengujian verifikasi data menggunakan data rekaman masker dan non-masker.

4. HASIL DAN PEMBAHASAN

Berdasarkan skenario pengujian yang telah dipaparkan, proses pengujian akan dijelaskan secara bertahap sesuai dengan urutan dalam sub bab skenario pengujian. Pengamatan hasil pengujian dalam skenario uji pertama dilakukan menggunakan nilai parameter awal ekstraksi fitur MFCC dengan panjang *frame* 20ms, dan 12 koefisien MFCC. Penggunaan nilai tersebut ditentukan berdasarkan nilai terkecil dari setiap parameter uji ekstraksi fitur MFCC. Kemudian hasil evaluasi ditentukan dengan mengamati hasil kombinasi performa pada nilai akurasi, presisi, dan *recall*.

4.1. Pengujian pengaruh *learning rate* dan jumlah epoch

Pengujian ini bertujuan untuk mengetahui pengaruh *learning rate* dan jumlah *epoch* dikarenakan nilai parameter tersebut tidak memiliki ketetapan sehingga dilakukan pengujian untuk mendapatkan model terbaik dari parameter tersebut. Pemilihan parameter *learning rate* didasarkan pada penelitian sebelumnya yang melihat perubahan signifikan kecepatan pembelajaran di antara nilai 0.001 dan 0.01 [30]. Pada penelitian ini, digunakan 3 variasi *learning rate* yaitu 0.1, 0.01 dan 0.001. Sedangkan untuk *epoch* digunakan 50, 100 dan 200. Hasil dari tiap variasi *learning rate* dan jumlah *epoch* disajikan pada Tabel II.

TABEL II. HASIL PENGUJIAN LEARNING RATE DAN EPOCH.

Learning Rate	*Epoch*	*Accuracy* (%)	*Precision* (%)	*Recall* (%)
0.1	50	81.98	76.97	91.76
	100	81.98	76.97	91.76
	200	81.98	76.97	91.76
0.01	50	82.57	77.56	92.16
	100	81.98	76.97	91.76
	200	81.78	76.72	91.76
0.001	50	82.57	77.74	91.76
	100	82.77	77.81	92.16
	200	82.97	77.89	92.55

Pada Tabel II dapat diketahui bahwa pada proses klasifikasi menggunakan LVQ hasil akurasi terbaik didapatkan pada pengujian *learning rate* 0.001 dan jumlah *epoch* 200 dengan nilai akurasi sebesar 82.97%, presisi sebesar 77.89% dan *recall* sebesar 92.55%. Kemudian pengujian parameter selanjutnya menggunakan arsitektur LVQ dengan nilai *learning rate* 0.001 dan jumlah *epoch* sebesar 200.

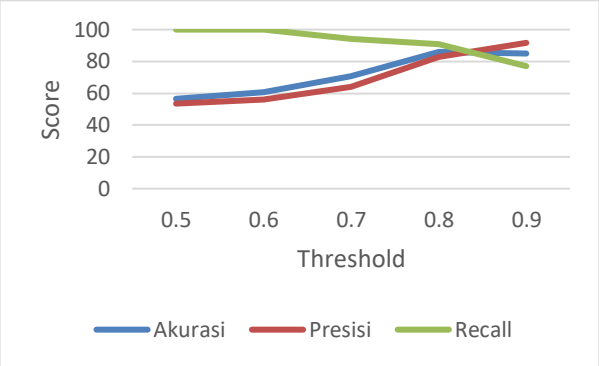
4.2. Pengujian pengaruh ambang batas *threshold*

Pengujian terhadap ambang batas *threshold* dilakukan untuk mengetahui batas minimal nilai untuk menentukan verifikasi suara pembicara. Pada pengujian ini, nilai variabel *threshold* berada pada rentang 0.5 – 0.9. Nilai tersebut digunakan untuk mendapatkan *term* yang memiliki nilai signifikan. Hasil dari pengujian pengaruh ambang batas *threshold* disajikan pada Tabel III

TABEL III. HASIL PENGUJIAN AMBANG BATAS THRESHOLD.

Threshold	*Accuracy* (%)	*Precision* (%)	*Recall* (%)
0.5	55.05	52.90	100
0.55	56.44	53.68	100
0.6	58.42	54.84	100
0.65	61.78	56.92	100
0.7	65.15	59.38	98.04
0.75	69.11	62.41	97.65
0.8	77.03	70.03	95.29
0.85	82.97	77.89	92.55
0.9	81.76	75.74	95.75

Berdasarkan Tabel III dapat diketahui bahwa ambang batas *threshold* terbaik didapatkan pada nilai 0.85 dengan hasil evaluasi yaitu akurasi sebesar 82.97%, presisi sebesar 77.89% dan *recall* sebesar 92.55%



Gambar 9. Pengaruh nilai *threshold* terhadap nilai evaluasi.

Diagram di atas menunjukkan, semakin meningkatnya nilai ambang batas *threshold* nilai akurasi dan presisi juga meningkat, namun untuk *recall* mengalami penurunan, nilai terbaik didapatkan di ambang batas *threshold* 0.85 yang dapat dilihat dari perpotongan kenaikan nilai akurasi, presisi dan *recall*.

4.3. Pengujian panjang *frame* dan koefisien MFCC

Beberapa pengujian sebelumnya dapat diambil hasil model LVQ yang menghasilkan hasil terbaik yaitu dengan menggunakan *learning rate* 0.001, 200 *epoch* dan 0.8 *threshold*. Pada penelitian ini, akan diujikan tiga variasi jumlah panjang *frame* (20ms, 25ms, 30ms) dan koefisien MFCC (12, 20 dan 40). Parameter yang digunakan berdasarkan hasil penelitian sebelumnya durasi jendela optimal ucapan didapatkan dalam rentang 15 – 35ms, dan untuk nilai koefisien MFCC 13, 20 dan 40 [31][32]. Hasil pengujian dapat dilihat pada Tabel IV.

TABEL IV. HASIL PENGUJIAN PANJANG FRAME DAN KOEFISIEN MFCC.

Panjang *Frame* (ms)	*Coeff MFCC*	*Accuracy* (%)	*Precision* (%)	*Recall* (%)
20	12	82.97	77.89	92.55
	20	89.31	87.08	92.55
	40	90.10	88.97	91.76
25	12	83.56	78.29	93.33
	20	89.50	87.13	92.94
	40	89.70	89.80	89.80
30	12	83.17	77.96	92.94
	20	89.70	87.18	93.33
	40	89.90	90.48	89.41

Berdasarkan hasil pengujian pada Tabel V, performa terbaik didapatkan pada panjang *frame* sebesar 20ms dengan jumlah koefisien MFCC sebanyak 40 fitur. Panjang *frame* memilki pengaruh dalam informasi yang dihasilkan dari ekstraksi MFCC, panjang *frame* yang kecil mempengaruhi kepadatan informasi yang ditangkap dan dalam pengujian ini 20ms mendapatkan panjang *frame* optimal [33].

Hasil pengujian jumlah koefisien MFCC didapatkan akurasi maksimum pada 40 ekstraksi fitur. Akurasi maksimum dengan 12 fitur yaitu 82.97%, dan 20 fitur mendapatkan 89.31%, hal ini menunjukkan semakin banyak fitur yang digunakan akurasi semakin membaik seperti yang didapatkan pada koefisien 40 yaitu 90.10%.

4.4. Pengujian Pengaruh *Similarity Score*

Selanjutnya dilakukan pengujian nilai *similarity score* pada data klaim (data uji pembicara asli) dan data bukan pembicara asli (peniru pembicara asli). Tujuan utamanya adalah untuk melihat batas minimal data tidak autentik antara data pembicara asli dan bukan pembicara asli. Perhitungan dilakukan menggunakan *cosine similarity* [21]. Hasil masing-masing dapat dilihat pada tabel 4.4 dan 4.5 seperti berikut

TABEL V. HASIL PENGUJIAN SIMILARITY SCORE DATA KLAIM

Kelas	Data klaim	Kelas	Data klaim
Ajem	81.72%	Indira	97.52%
Arsan	99.43%	Iyan	95.13%
Bilya	98.82%	May	93.69%
Bintang	91.67%	Moren	95.12%
Caca	97.22%	Nadya	79.01%
Cici	96.85%	Odik	99.79%
Dela	98.51%	Pangeran	96.98%
Desi	95.54%	Putami	88.16%
Ega	98.27%	Ridho	96.85%
Faza	95.21%	Risman	93.69%
Fira	87.33%	Safira	91.45%
Ifa	97.35%	Sulhan	99.10%
Ikii	84.53%	**Rata-rata**	**93.96%**

Hasil pada tabel V mendapatkan rata-rata nilai similarity score sebesar 93.96%. Hasil tersebut menyatakan bahwa data klaim memiliki kedekatan data yang tinggi dengan data pembicara asli. Kedekatan data ini menunjukkan suara pembicara di data latih di rekam oleh pembicara yang sama di data klaim, meskipun dengan perekaman di waktu dan tempat yang berbeda

TABEL VI. HASIL PENGUJIAN SIMILARITY SCORE DATA BUKAN PEMBICARA

Kelas	Data Peniru	Kelas	Data Peniru
Ajem	61.26%	Indira	84.29%
Arsan	81.45%	Iyan	66.53%
Bilya	77.44%	May	54.45%
Bintang	61.11%	Moren	77.59%
Caca	74.02%	Nadya	19.11%
Cici	67.87%	Odik	81.63%
Dela	73.62%	Pangeran	64.73%
Desi	70.36%	Putami	83.15%
Ega	69.68%	Ridho	72.28%
Faza	70.84%	Risman	70.26%
Fira	22.72%	Safira	75.00%
Ifa	73.67%	Sulhan	82.35%
Ikii	54.13%	**Rata-rata**	**67.58%**

Berdasarkan Tabel VI diketahui bahwa bobot LVQ memiliki kedekatan data dengan data bukan pembicara *similarity score* sebesar 67.58%. Hasil tersebut menyatakan bahwa data uji bukan pembicara memiliki kedekatan data yang rendah dengan data pembicara asli. Nilai *similarity score* yang rendah menunjukkan suara manusia memiliki perbedaan antara satu individu dengan individu lain [5].

Nilai masing-masing fitur pada masing-masing kelas memiliki perbedaan persebaran data yang signifikan yaitu pada data pembicara dengan data peniru. Hal tersebut menunjukkan bahwa suara memiliki nilai yang berbeda satu dengan yang lain. Acuan nilai minimal pembeda dalam penelitian ini dapat dihitung margin rata-rata dari data pembicara dan peniru yaitu sebesar 80%.

4.5. Pengujian verifikasi data masker dan *non*-masker

Pengujian sebelumnya menghasilkan nilai terbaik didapatkan pada 15 *training* data. Selanjutnya dilakukan pengujian menggunakan dua data klaim yang berbeda yaitu pengujian data masker dan *non*-masker. Berikut adalah hasil pengujian pada data masker dan *non*-masker.

TABEL VII. HASIL PENGUJIAN DATA MASKER DAN NON-MASKER.

Data klaim	*Accuracy* (%)	*Precision* (%)	*Recall* (%)
Masker	86.2	87.86	84
Non-masker	90	88.97	91.17

Berdasarkan hasil pengujian pada Tabel 4.6 dapat disimpulkan bahwa model tidak dapat mengenali pengguna saat menggunakan masker dengan baik. Hal tersebut dapat dilihat pada tabel bahwa nilai akurasi menurun menjadi 86.2%, presisi 87.86% dan *recall* 84%.

TABEL VIII. CONFUSSION MATRIX PENGUJIAN DATA MASKER.

		Kelas Aktual	
		Positif	Negatif
Prediksi	*True*	210	40
	False	29	221

TABEL IX. CONFUSSION MATRIX PENGUJIAN DATA NON - MASKER.

		Kelas Aktual	
		Positif	Negatif
Prediksi	*True*	234	21
	False	29	221

Berdasarkan Tabel VIII dan Tabel IX dapat dilihat terjadinya peningkatan signifikan yaitu pada *true negative* didapatkan perubahan dari 21 menjadi 40, hal tersebut dapat terjadi karena adanya perbedaan suara yang keluar karena penggunaan masker ketika melakukan perekaman.

Akurasi yang berbeda dari data masker dan non masker disebabkan karena fase sinyal mungkin tidak mengalami pergeseran yang sangat besar ketika menggunakan masker sehingga saat menggunakan masker memiliki klasifikasi yang rendah [6]. Peningkatan akurasi pengenalan dapat ditingkatkan dengan menggunakan *low level descriptor* untuk mendapatkan ekstraksi amplitudo atau sinyal, sehingga dapat melihat nilai dari variasi suara seperti *zero crossing rate*, *chrome*, *tonnettz* [1].

5. KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan penelitian yang telah dilakukan terdapat beberapa hal yang dapat disimpulkan antara lain sebagai berikut.

1. Model terbaik yang dihasilkan pada penelitian ini yaitu dengan menggunakan *learning rate* 0.001, 200 *epoch* dengan data *training* yang memiliki panjang *frame* 30ms, 20 koefisien MFCC dan ambang batas *threshold* sebesar 0.85 dengan akurasi sebesar 90%.
2. Hasil pengujian menggunakan data masker dan non-masker masing-masing mendapatkan akurasi sebesar 86.2% dan 90%.
3. Nilai standar minimal data tidak valid didapatkan rata – rata margin *similarity score* sebesar 80% berdasarkan nilai *similarity score* pada data klaim sebesar 93.96% dan data bukan pembicara sebesar 67.58%.

5.2. Saran

Penelitian yang telah dilakukan masih jauh dari kata sempurna, adapun beberapa saran yang dapat penulis berikan untuk pengembangan penelitian ini antara lain sebagai berikut.

1. Menambahkan jumlah *dataset* untuk masing-masing pembicara agar *classifier* dapat mengenali lebih banyak pembicara.
2. Menambahkan perhitungan metode statistik seperti *Gaussian Mixture Model* (GMM) atau *Discrete Wavelet Transform* (DWT) sebagai masukan model.
3. Mencoba menggunakan *low level descriptor* seperti *loudness, energi, zero crossing rate* pada proses *pre-processing* sinyal suara untuk meningkatkan pengenalan suara menggunakan medium masker

UCAPAN TERIMA KASIH

Ucapan terima kasih diberikan kepada partisipan yang telah bersedia membantu pembuatan rekaman *dataset* menggunakan ponsel pribadi masing-masing.

DAFTAR PUSTAKA

[1] T. Roy, T. Marwala, and S. Chakraverty, “Speech Emotion Recognition Using Neural Network and Wavelet Features,” *Lect. Notes Mech. Eng.*, vol. 10, no. 4, pp. 427–438, 2020, doi: 10.1007/978-981-15-0287-3_30.

[2] P. E. Zulfikar, S. H. Sitorus, U. Ristian, J. Rekayasa, and S. Komputer, “Sistem Presensi Menggunakan Verifikasi Palm Print Dengan Metode Principal Component Analysis Dan Euclidean Distance,” *Coding J. Komput. dan Apl.*, vol. 09, no. 01, pp. 33–43, 2021.

[3] T. M. Tamtelahitu, “Perancangan Sistem Absensi Pintar Mahasiswa Menggunakan Teknik Qr Code Dan Geolocation,” *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 6, no. 1, pp. 114–125, 2021, doi: 10.29100/jipi.v6i1.1894.

[4] H. Rohman, U. Darussalam, and N. D. Natashia,

"Sistem Presensi Fingerprint Berbasis Smartphone Android," *J I M P - J. Inform. Merdeka Pasuruan*, vol. 5, no. 1, pp. 1–5, 2020, doi: 10.37438/jimp.v5i1.241.

[5] E. R. Sharipova, A. A. Horoshiy, and N. A. Kotlyarov, "Student Voice Identification Method," *Proc. 2021 IEEE Conf. Russ. Young Res. Electr. Electron. Eng. ElConRus 2021*, no. 2, pp. 647–651, 2021, doi: 10.1109/ElConRus51938.2021.9396443.

[6] R. K. Das and H. Li, "Classification of Speech with and without Face Mask using Acoustic Features," *2020 Asia-Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. APSIPA ASC 2020 - Proc.*, pp. 747–752, 2020.

[7] A. Al-Qaisi, "Arabic word dependent speaker identification system using artificial neural network," *Int. J. Circuits, Syst. Signal Process.*, vol. 14, no. July, pp. 290–295, 2020, doi: 10.46300/9106.2020.14.41.

[8] S. M. Widodo, E. Siswanto, and O. Sudjana, "Penerapan Metode Mel Frequency Ceptral Coefficient dan Learning Vector Quantization untuk Text-Dependent Speaker Identification," *Sukoreno Mukti Widodo, Elisafina Siswanto, Eotomo Sudjana*, vol. 11, no. 1, pp. 15–20, 2016.

[9] V. Indrawati and Y. Gunawan, "Penggunaan Algoritma Learning Vector Quantization Dalam Mengenali Suara Manusia Untuk Kendali Quadrotor," vol. 2014, no. Sentika, 2014.

[10] A. Gustanto and I. Susilawati, "Pengenalan Suara Untuk Identifikasi Personal Menggunakan LVQ Voice Recognition For Personal Identification Using LVQ," no. 84, pp. 9–17, 2018.

[11] F. Paath, L. A. Latumakulita, C. Montolalu, and Y. Langi, "Pengenalan Suara Manusia Menggunakan Convolutional Neural Network Studi Kasus Suara Dosen Program Studi Sistem Informasi Universitas Sam Ratulangi," *Konf. Nas. Ilmu Komput.*, pp. 215–218, 2021.

[12] A. Maurya, D. Kumar, and R. K. Agarwal, "Speaker Recognition for Hindi Speech Signal using MFCC-GMM Approach," *Procedia Comput. Sci.*, vol. 125, pp. 880–887, 2018, doi: 10.1016/j.procs.2017.12.112.

[13] Y. Miftahuddin, M. B. Amin, and R. Budiraharjo, "Implementation of MFCC And LVQ Methods For Learning English Pronunciation," in *International Conference on Green Technology and Design*, 2020, pp. 123–129.

[14] P. Mishra and P. K. Mishra, "Text-Dependent Multilingual Speaker Identification using Learning Vector Quantization and PSO-GA Hybrid Model," *Int. Res. J. Eng. Technol.*, vol. 03, no. 09, pp. 1034–1040, 2016.

[15] P. Prasetyawan, "Perbandingan Identifikasi Pembicara Menggunakan Mfcc Dan Sbc Dengan Ciri Pencocokan Lbg-Vq," vol. 2016, no. Sentika, pp. 18–19, 2018, doi: 10.31227/osf.io/85k9u.

[16] A. Kurniawan, "Verifikasi Suara menggunakan Jaringan Syaraf Tiruan dan Ekstraksi Ciri Mel Frequency Cepstral Coefficient," *J. Sist. Inf. Bisnis*, vol. 7, no. 1, p. 32, 2017, doi: 10.21456/vol7iss1pp32-38.

[17] H. Beigi, *Fundamentals of Speaker Recognition*. 2011.

[18] S. Jothilakshmi and V. N. Gudivada, *Large Scale Data Enabled Evolution of Spoken Language Research and Applications*, 1st ed., vol. 35. Elsevier B.V., 2016.

[19] H. Santoso and M. Fakhriza, "Perancangan Aplikasi Keamanan File Audio Format Wav (Waveform) Menggunakan Algoritma Rsa," *Algoritm. J. Ilmu Komput. dan Inform.*, vol. 2, no. 1, pp. 47–54, 2018, [Online]. Available: http://jurnal.uinsu.ac.id/index.php/algoritma/article/view/1615.

[20] K. Aizat, O. Mohamed, M. Orken, A. Ainur, and B. Zhumazhanov, "Identification and authentication of user voice using DNN features and i-vector," *Cogent Eng.*, vol. 7, no. 1, 2020, doi: 10.1080/23311916.2020.1751557.

[21] S. Hourri and J. Kharroubi, "A Novel Scoring Method Based on Distance Calculation for Similarity Measurement in Text-Independent Speaker Verification," *Procedia Comput. Sci.*, vol. 148, pp. 256–265, 2019, doi: 10.1016/j.procs.2019.01.068.

[22] M. Sigmund, "Automatic Speaker Recognition by Speech Signal," *Front. Robot. Autom. Control*, no. May, 2008, doi: 10.5772/6333.

[23] L. Fausett, *Fundamental of Neural Networks: Architectures, Algorithms, and Applications*. Englewood Cliffs: NJ: Prentice-Hall, 1994.

[24] C. Hadi and M. R. Ma'arif, "Implementasi Cosine Similarity Dalam Aplikasi Pencarian Ayat Al-Qur'an Berbasis Android," *Foreign Aff.*, vol. 6, no. 2, pp. 70–79, 2017.

[25] J. Wang *et al.*, "Systematic analysis and prediction of type IV secreted effector proteins by machine learning approaches," *Brief. Bioinform.*, vol. 20, no. 3, pp. 931–951, 2017, doi: 10.1093/bib/bbx164.

[26] S. Soleymani, A. Dabouei, S. M. Iranmanesh, H. Kazemi, J. Dawson, and N. M. Nasrabadi, "Prosodic-enhanced siamese convolutional neural networks for cross-device text-independent speaker verification," *2018 IEEE 9th Int. Conf. Biometrics Theory, Appl. Syst. BTAS 2018*, pp. 1–7, 2018, doi:

10.1109/BTAS.2018.8698585.

[27] C. Kumar, F. Ur Rehman, S. Kumar, A. Mehmood, and G. Shabir, "Analysis of MFCC and BFCC in a speaker identification system," *2018 Int. Conf. Comput. Math. Eng. Technol. Inven. Innov. Integr. Socioecon. Dev. iCoMET 2018 - Proc.*, vol. 2018-Janua, no. December, pp. 1–5, 2018, doi: 10.1109/ICOMET.2018.8346330.

[28] N. Uddin *et al.*, "Development of Voice Recognition for Student Attendance," *Glob. J. Hum. Soc. Sci. G Linguist. Educ.*, vol. 16, no. 1, pp. 1–8, 2016.

[29] D. Anggraeni, W. S. M. Sanjaya, M. Y. S. Nurasyidiek, and M. Munawwaroh, "The Implementation of Speech Recognition using Mel-Frequency Cepstrum Coefficients (MFCC) and Support Vector Machine (SVM) method based on Python to Control Robot Arm," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 288, no. 1, 2018, doi: 10.1088/1757-899X/288/1/012042.

[30] L. N. Smith, "Cyclical learning rates for training neural networks," *Proc. - 2017 IEEE Winter Conf. Appl. Comput. Vision, WACV 2017*, no. April, pp. 464–472, 2017, doi: 10.1109/WACV.2017.58.

[31] S. Nainan and V. Kulkarni, "Enhancement in speaker recognition for optimized speech features using GMM, SVM and 1-D CNN," *Int. J. Speech Technol.*, vol. 24, no. 4, pp. 809–822, 2021, doi: 10.1007/s10772-020-09771-2.

[32] J. G. L. and K. K. W. Kuldip K. Paliwal, "Preference for 20-40 ms window duration in speech analysis Kuldip K . Paliwal , James G . Lyons and Kamil K . W ´ ojcicki Signal Processing Laboratory," *J. Theaudio Eng. Soc.*, pp. 154–162, 2010.

[33] F. D. Adhinata, D. P. Rakhmadani, and A. J. T. Segara, "Pengenalan Jenis Kelamin Manusia Berbasis Suara Menggunakan MFCC dan GMM," *Jorunal data Sci. IoT, Mach. Learn. Informatics*, vol. 1, no. 1, pp. 11–12, 2021.