**SMARTICS Journal**, Vol.12 No.2, 20206. p96-106
ISSN online: 2476-9754, ISSN print: 2623-0429
DOI : https://doi.org/10.21067/smartics.v12i2.14728


# Autentikasi Biometrik Suara menggunakan ECAPA-TDNN untuk Peningkatan Keamanan Sistem Kendali Pagar Rumah

**Ajib Hanani[a*], Mokhamad Amin Hariyadi[b], Johan Ericka Wahyu Prakasa[c], Shoffin Nahwa Utama[d]**
[a,b,c,d]UIN Maulana Malik Ibrahim Malang, Indonesia
[*]correspondence email: ajib@uin-malang.ac.id

***Abstract*—Conventional home gate security systems relying on physical keys are vulnerable to loss and theft, whereas standard voice-controlled smart home systems focus solely on speech recognition without verifying the speaker's identity. This study aims to integrate voice biometric authentication into an IoT-based home gate control system to deliver dual-layer security. An experimental design utilizing a prototyping methodology was employed. The system implements the ECAPA-TDNN Deep Learning algorithm on a Raspberry Pi edge computing device to verify the homeowner's voice identity using Cosine Similarity, alongside an ESP32 microcontroller to drive a stepper motor via MQTT. Findings indicate high biometric accuracy with a False Acceptance Rate (FAR) of 3.75%, a False Rejection Rate (FRR) of 2.08%, and an Equal Error Rate (EER) of 2.92%, while all functional mechanical controls and live CCTV streaming executed reliably in real-time. Combining ECAPA-TDNN with IoT effectively establishes a secure, responsive gate control system independent of third-party cloud services. Future research is suggested to evaluate system performance in multi-user settings with diverse family voice profiles.***

***Keywords*— ECAPA-TDNN; Gate Control System; Internet of Things; Speaker Verification; Voice Biometrics**

***Abstrak*–Sistem keamanan pagar rumah konvensional berbasis kunci fisik rentan terhadap risiko kehilangan dan pencurian, sedangkan sistem smart home berbasis perintah suara umum hanya berfokus pada pengenalan ucapan (speech recognition) tanpa memverifikasi identitas pembicara. Penelitian ini bertujuan mengintegrasikan autentikasi biometrik suara ke dalam sistem kendali pagar rumah berbasis IoT untuk menghadirkan keamanan ganda. Metode yang digunakan adalah eksperimental berbasis prototyping. Sistem menerapkan algoritma Deep Learning ECAPA-TDNN pada perangkat edge computing Raspberry Pi untuk memverifikasi identitas suara pemilik rumah menggunakan Cosine Similarity, serta mikrokontroler ESP32 sebagai penggerak motor stepper via protokol MQTT. Temuan penelitian menunjukkan kinerja biometrik suara yang akurat dengan nilai False Acceptance Rate (FAR) 3,75%, False Rejection Rate (FRR) 2,08%, dan Equal Error Rate (EER) 2,92%, serta seluruh fungsi kontrol mekanis dan pemantauan CCTV berjalan valid secara real-time. Integrasi ECAPA-TDNN dan IoT berhasil menciptakan sistem kendali pagar rumah yang aman, responsif, dan independen dari layanan cloud pihak ketiga. Penelitian selanjutnya disarankan untuk menguji keandalan sistem pada skala multi-user menggunakan variasi profil suara anggota keluarga.***

***Kata Kunci*— Biometrik Suara; ECAPA-TDNN; Internet of Thing; Sistem Kendali Pagar; Speaker Verification**

## I. PENDAHULUAN

Perkembangan teknologi *Internet of Things* (IoT) telah mengubah paradigma keamanan dari sistem konvensional menuju *smart home*. Salah satu aspek krusial dalam keamanan rumah adalah sistem kendali akses pintu pagar. Secara tradisional, akses pintu pagar bergantung pada kunci fisik yang rentan terhadap risiko kehilangan, kerusakan, duplikasi, kelupaan atau dicuri [1], [11].

Sebagai solusi, berbagai penelitian telah mengembangkan sistem kendali pagar/pintu berbasis IoT yang memanfaatkan aplikasi *smartphone* [18] hingga perintah suara [3], [5], [7], [13], [15], [17], [19], [24], [26]. Konsep dasar dari sistem ini adalah memberikan kemudahan akses tanpa sentuhan, yang sangat bermanfaat bagi penyandang disabilitas atau lansia [2], [21]. Namun, mayoritas sistem yang ada saat ini hanya berfokus pada *speech recognition*, yaitu sistem yang memahami "apa yang diucapkan", tetapi tidak memverifikasi "siapa yang mengucapkannya" [11].

Teknologi dalam pemrosesan sinyal suara untuk mengidentifikasi identitas pembicara dikenal sebagai *speaker verification*. Evolusi teknologi ini telah berkembang pesat dari model statistik *i-vector* menuju pendekatan *Deep Learning*. Hingga beberapa tahun terakhir, muncul arsitektur *x-vector* yang berbasis *Time Delay Neural Network* (TDNN).

Terobosan terbaru yang menjadi *State of the Art* (SOTA) saat ini adalah arsitektur ECAPA-TDNN (*Emphasized Channel Attention, Propagation and Aggregation in Time Delay Neural Network*). Desplanques pada tahun 2020 pertama kali memperkenalkan arsitektur ECAPA-TDNN yang memperbaiki kelemahan *x-vector* standar. Penelitian ini membuktikan bahwa dengan menambahkan *Emphasized Channel Attention, Propagation and Aggregation (ECAPA)*, sistem mampu menangkap fitur suara pembicara dengan lebih detail [9]. Hasil studi komparasi yang membandingkan berbagai model *deep speaker embedding* populer seperti WavLM, TitaNet, ECAPA-TDNN dan PyAnnote untuk *speaker verification* menunjukkan bahwa ECAPA-TDNN memiliki *Equal Error Rates* (EER) yang paling rendah, yaitu 1,71% [6]. Dalam aspek ketahanan (*robustness*), He pada tahun 2024 mengembangkan varian sistem berbasis TDNN untuk lingkungan bising dan menemukan bahwa arsitektur dasar ECAPA-TDNN memiliki toleransi yang baik terhadap *noise* [10]. Hal ini relevan dengan kondisi pagar rumah yang berada di luar ruangan. Selain itu, Sigona & Grimaldi pada tahun 2024 memvalidasi penggunaan ECAPA-TDNN untuk keperluan forensik, menegaskan bahwa algoritma ini cukup akurat untuk membedakan identitas individu dalam kondisi kasus nyata, bukan hanya data sintetik [23]. Selain itu, ECAPA-TDNN juga memiliki tingkat akurasi 90,51% dalam mendeteksi patologi suara secara otomatis [22] dan memiliki tingkat akurasi 85% dalam mendeteksi suara kelelahan [4]. Oleh karena itu, penelitian ini memilih algoritma ECAPA-TDNN sebagai solusi.

Berdasarkan tinjauan literatur terhadap artikel ilmiah yang relevan, ditemukan *research gap* (kesenjangan penelitian) yang signifikan**.** *Research gap* pertama adalah belum terintegrasinya algoritma *High-End* ke IoT. Meskipun ECAPA-TDNN telah terbukti unggul dalam *speaker verification* [6], [9] dan bahkan digunakan untuk keperluan forensik dalam kasus kriminal [16], penerapannya secara spesifik pada sistem kendali pagar rumah (*smart gate*) masih belum ada. Kebanyakan implementasi ECAPA-TDNN masih berada pada tataran eksperimen *dataset* publik seperti VoxCeleb [25] atau *speaker diarization* [8], dan belum banyak diaplikasikan secara *embedded* untuk sistem keamanan rumah. *Research gap* kedua adalah keterbatasan keamanan pada implementasi IoT saat ini. Mayoritas penelitian sistem kendali pagar masih mengandalkan *Google Assistant* yang hanya menerjemahkan suara ke teks [3], [13], [15], [17], [19], [24]. Kelemahan fatal dari pendekatan ini adalah siapa pun yang mengetahui kata kunci perintah dapat membuka akses, sehingga aspek autentikasi biometrik terabaikan. *Research gap* ketiga adalah ketergantungan pada *cloud* pihak ketiga (seperti *Google Assistant*) yang memiliki isu privasi data.

Terdapat *novelty* (kebaruan) yang ditawarkan dalam penelitian ini. *Novelty* yang pertama adalah mengintegrasikan algoritma *High-End ke IoT*. Membawa algoritma ECAPA-TDNN, yang biasanya hanya digunakan di ranah riset sinyal digital, dimasukkan ke dalam implementasi praktis *Smart Gate System*. *Novelty* yang kedua adalah peningkatan keamanan ganda. Menggabungkan kemudahan perintah suara dengan keamanan biometrik, dimana sistem hanya akan merespons jika perintah suara yang benar diucapkan oleh pemilik suara yang terdaftar (menggabungkan *Speech Recognition* dan *Speaker Verification*), mengatasi kelemahan sistem berbasis *Google Assistant* yang umum digunakan. *Novelty* yang ketiga adalah kemandirian sistem. Mengurangi ketergantungan pada *cloud* pihak ketiga (seperti *Google Assistant*) dengan memproses inferensi model secara lokal atau *edge-server*, meningkatkan privasi dan responsivitas.

Penelitian ini mendesak dilakukan karena meningkatnya risiko keamanan pada sistem *smart home* yang hanya mengandalkan konektivitas tanpa autentikasi biometrik [11]. Selain itu, kebutuhan akan sistem keamanan yang tidak hanya cerdas (*smart*) tetapi juga aman (*secure*) dan akurat dalam membedakan pemilik rumah dengan penyusup, terutama di tengah meningkatnya adopsi teknologi IoT di masyarakat [12].

Penelitian ini diharapkan dapat memberikan kontribusi signifikan, baik secara teoritis maupun praktis. Kontribusi secara teoritis dalam penelitian ini yaitu memberikan bukti empiris tentang efektivitas algoritma ECAPA-TDNN ketika diterapkan pada perangkat keras yang terbatas dan untuk menambah khazanah keilmuan tentang *TinyML* atau *Edge AI*, khususnya dalam penerapan model berat seperti TDNN pada arsitektur sistem keamanan rumah. Sedangkan kontribusi secara praktis dalam penelitian ini yaitu memberikan solusi alternatif bagi masyarakat untuk meningkatkan sistem keamanan rumah, membantu lansia atau penyandang disabilitas daksa yang kesulitan menggunakan kunci fisik dan mengurangi ketergantungan pada layanan *cloud* pihak ketiga (seperti *Google Assistant*) yang memiliki isu privasi data.

## II. METODE PENELITIAN

Penelitian ini menggunakan pendekatan eksperimental dengan menerapkan metode *prototyping* [14].

Metode ini dipilih karena memungkinkan pengembangan sistem secara terstruktur dan fleksibel melalui beberapa tahapan yang sangat cocok untuk integrasi teknologi *Internet of Things* (IoT) dengan *Artificial Intelligence* (AI) berbasis *Edge Computing*. Secara garis besar, alur pengembangan prototipe dibagi menjadi 5 tahapan utama, yaitu: (1) Identifikasi Kebutuhan Sistem, (2) Perancangan dan Pembangunan Prototipe, (3) Evaluasi Prototipe, (4) Perbaikan Prototipe, dan (5) Implementasi Sistem Akhir.

*A. Identifikasi Kebutuhan Sistem*

Tahap awal berfokus pada analisis kebutuhan *hardware*, *software* serta spesifikasi data audio yang digunakan dalam membangun *Smart Voice Gate System*.

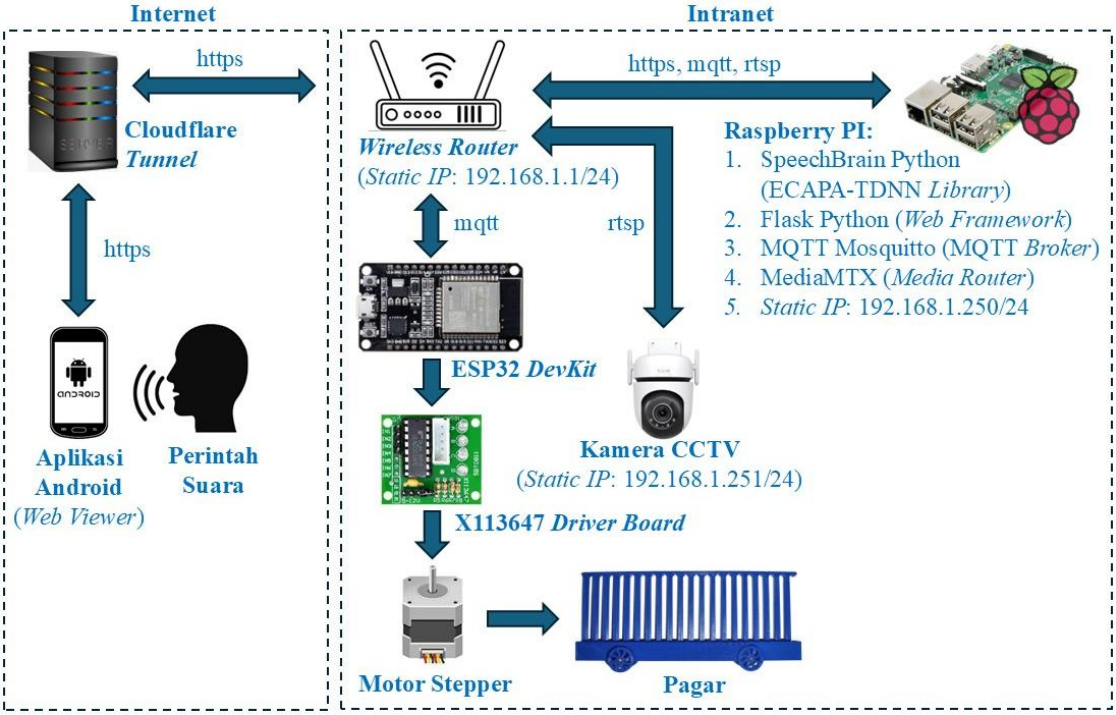
- Perangkat Keras (*Hardware*) yang digunakan adalah *Single Board Computer* Raspberry Pi, Mikrokontroler ESP32 DevKit, X113647 *Driver Board*, *Motor Stepper*, Kamera CCTV, *Wireless Router*, serta miniatur pagar.
- Perangkat Lunak (*Software*) yang digunakan adalah Python dengan pustaka SpeechBrain, *web framework* Flask, MQTT *Broker* Mosquitto, MediaMTX, Cloudflare *Tunnel*, Google *Speech Recognizer*, dan aplikasi Android yang dibangun via MIT App Inventor.
- Spesifikasi Sinyal Audio: Format WAV dengan spesifikasi saluran audio *mono channel, sample rate* (laju sampel) 16kHz, format *encoding* PCM 16-bit dan 10 sampel suara *enrollment* untuk tiap kata kunci "buka" maupun "tutup" pagar.

*B. Perancangan dan Pembangunan Prototipe*

Pada tahap ini, prototipe dirancang dan dibangun ke dalam arsitektur sistem IoT serta logika pemrosesan berbasis *Input-Process-Output* (IPO).

- Desain *Smart Voice Gate System*

Sistem dirancang dengan konsep *Edge Computing* dimana pemrosesan berat (*AI Inference*) dilakukan pada *Single Board Computer* Raspberry Pi, sedangkan kontrol aktuator dilakukan oleh mikrokontroler ESP32 *DevKit* seperti yang disajikan pada Gambar 1 di bawah ini.



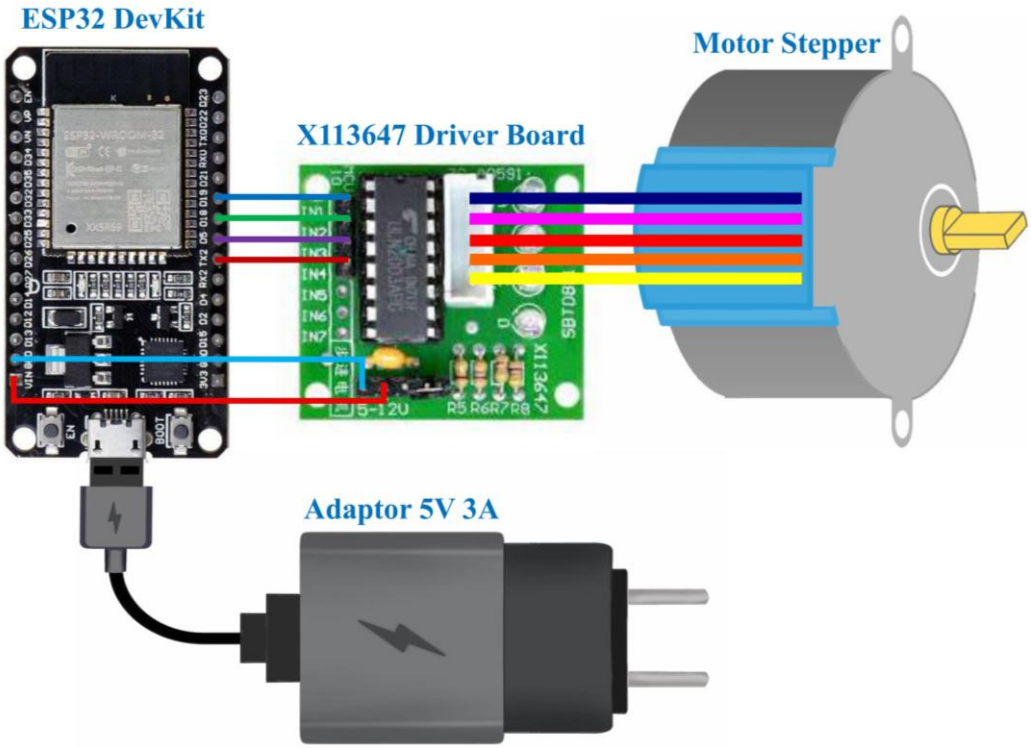

Gambar 1. Desain *Smart Voice Gate System*

Penjelasan desain *Smart Voice Gate System* sebagai berikut. Perangkat Raspberry Pi, *Wireless Router*, ESP32 *DevKit,* Modul Driver X113647 (berbasis chip ULN2003A), *Motor Stepper* dan miniatur pagar berada dalam satu jaringan lokal/intranet. Raspberry Pi sebagai *Inference Engine* ECAPA-TDNN, server aplikasi berbasis *web,* MQTT *Broker*, *Media Router* dan memiliki *Static IP Address*. *Library* SpeechBrain Python digunakan untuk mendukung ECAPA-TDNN dalam pemrosesan suara [20]. Aplikasi *Smart Voice Gate System* berbasis web dibangun dengan menggunakan Flask Python. Aplikasi berbasis Android juga

dibangun menggunakan *web viewer component* di MIT App Inventor untuk menampilkan aplikasi berbasis web tersebut. MQTT Mosquitto di-*install* dan dikonfigurasi sebagai MQTT *Broker*. MediaMTX di-*install* dan dikonfigurasi sebagai *Media Router*. Aplikasi berjalan di jaringan lokal/intranet dan dapat diakses secara publik/internet secara aman (*https*) menggunakan Cloudflare *Tunnel*. *Tunnel* merupakan metode enkapsulasi data yang menciptakan koneksi aman untuk mentransmisikan data dari jaringan lokal/intranet ke jaringan publik/internet. *Wireless router* sebagai jembatan komunikasi antar perangkat dalam jaringan lokal/intranet dan memiliki *Static IP Address*. Pada *Smart Voice Gate System* ini, pengguna memberi perintah suara untuk membuka/menutup pagar rumah menggunakan aplikasi berbasis android maupun aplikasi berbasis web. Perintah suara pengguna dijadikan audio untuk diverifikasi dan dijadikan teks di Raspberry Pi. Jika perintah suara diverifikasi benar diucapkan oleh pemilik rumah yang terdaftar dan teks terdapat kata kunci buka/tutup pagar, maka hasil keputusan pagar dibuka/ditutup dikirim ke ESP32 melalui protokol MQTT. ESP32 *subscribe* ke MQTT *Broker* Raspberry Pi dan terhubung ke Modul Driver X113647 melalui kabel *jumper*. *Motor Stepper* sebagai aktuator penggerak untuk membuka/menutup miniatur pagar. Pengguna juga dapat melihat kamera CCTV melalui aplikasi tersebut untuk mengetahui kondisi pagar dalam kondisi terbuka/tertutup dan untuk mengetahui siapa orang yang akan masuk/keluar rumah. Kamera CCTV memiliki *Static IP Address.* Kamera CCTV dapat diakses menggunakan protokol RTSP dan dikonversi ke protokol LL-HLS (*Low-Latency HTTP Live Streaming*) menggunakan MediaMTX untuk mempercepat pengiriman video secara langsung melalui intranet maupun internet.

- Rangkaian Kabel (*Wiring*) Perangkat

Sebagai pelengkap desain *Smart Voice Gate System*, berikut disajikan Gambar 2 tentang rangkaian kabel (*wiring*) Adaptor, ESP32 DevKit, Driver X113647 dan *Motor Stepper.*

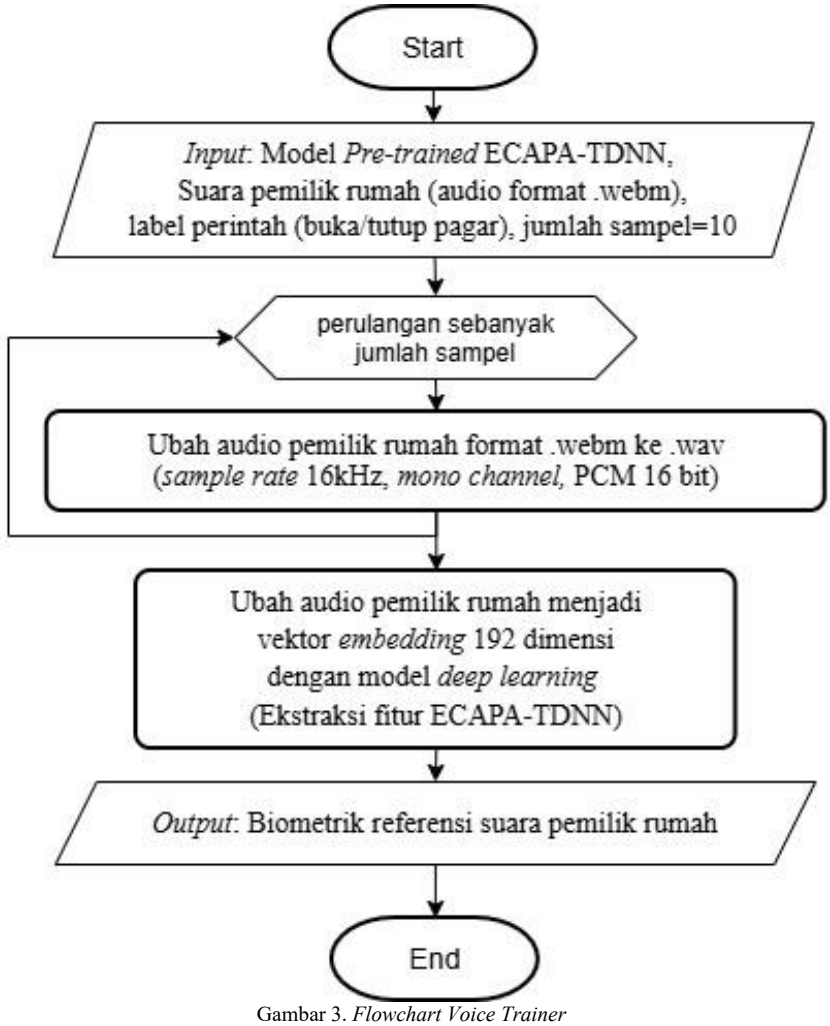



Gambar 2. Rangkaian kabel (*wiring*) Adaptor, ESP32 DevKit, Driver X113647 dan *Motor Stepper*

Penjelasan rangkaian kabel (*wiring*) Adaptor, ESP32 DevKit, Driver X113647 dan *Motor Stepper* adalah sebagai berikut. Adaptor 5V 3A menyuplai tegangan DC ke ESP32 DevKit, Driver X113647 dan *Motor Stepper*. Driver X113647 dihubungkan ke ESP32 menggunakan kabel *jumper* sebagai berikut: IN1 ke GPIO 19, IN2 ke GPIO 18, IN3 ke GPIO 5, IN4 ke GPIO 17, VCC (+) pada Driver X113647 ke Pin VIN (5V) ESP32, GND (-) pada Driver X113647 ke Pin GND ESP32. *Motor Stepper* dihubungkan langsung ke soket *output* 5-pin pada papan Driver X113647.

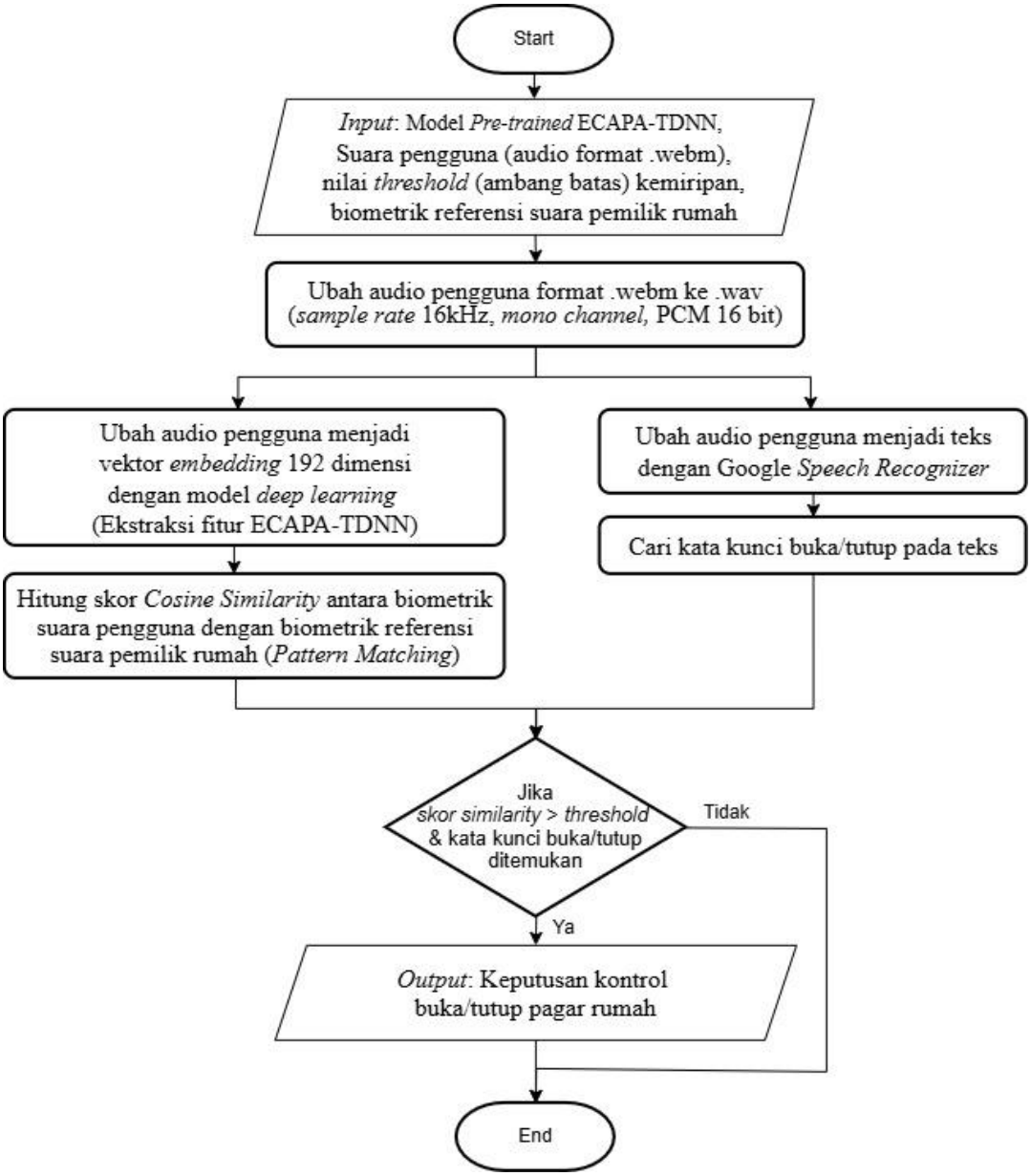
- Alur Logika Sistem *Input-Process-Output* (IPO)

Alur logika sistem dalam penelitian ini menggunakan pendekatan ***Input-Process-Output*** **(IPO)** untuk menggambarkan alur logika pemecahan masalah yang dituangkan dalam bentuk *flowchart voice trainer* dan *flowchart smart voice gate system* seperti yang disajikan pada gambar 3 dan gambar 4 sebagai berikut.




Gambar 3. *Flowchart Voice Trainer*

Penjelasan *flowchart voice trainer* sebagai berikut. *Input* berupa Model *Pre-trained* ECAPA-TDNN, perintah suara pemilik rumah dengan kata kunci "buka" atau “tutup” pagar dan jumlah sampel 10 untuk tiap kata kunci. Model *Pre-trained* ECAPA-TDNN dimuat ke dalam memori saat aplikasi pertama kali dijalankan. Perintah suara pemilik rumah melalui mikrofon berupa audio dengan *type data* “.webm” dikonversi menjadi “.wav” dengan spesifikasi saluran audio *mono channel, sample rate* (laju sampel) 16kHz serta format *encoding* PCM 16-bit yang diulang sebanyak 10 kali untuk tiap kata kunci "buka" maupun “tutup” pagar. Untuk kata kunci “buka” pagar terdapat sebanyak 10 *file* dan untuk kata kunci “tutup” pagar terdapat sebanyak 10 *file*. Selanjutnya tiap audio dilakukan ekstraksi fitur (ECAPA-TDNN) yang merupakan proses inti dimana model *deep learning* mengubah audio pemilik rumah menjadi vektor *embedding* 192-dimensi. Fitur *Channel Attention* memilah informasi penting dari *noise*. Nilai rata-rata ke-10 vektor *embedding* dihitung untuk menghasilkan *anchor embedding* yang merepresentasikan karakteristik khas pita suara. *Output* dari proses tersebut berupa 2 *anchor embedding* dengan *type data* “.pt” yang dijadikan sebagai biometrik referensi suara pemilik rumah, dimana 1 biometrik referensi suara untuk buka pagar dan 1 biometrik referensi suara untuk tutup pagar.




Gambar 4. *Flowchart Smart Voice Gate System*

Penjelasan *flowchart smart voice gate system* sebagai berikut. *Input* berupa Model *Pre-trained* ECAPA-TDNN, perintah suara pengguna (pemilik rumah dan orang asing) dengan kata kunci "buka" atau “tutup” pagar, nilai *threshold* (ambang batas) kemiripan dan biometrik referensi suara pemilik rumah. Model *Pre-trained* ECAPA-TDNN dimuat ke dalam memori saat aplikasi pertama kali dijalankan. Perintah suara pengguna melalui mikrofon berupa audio dengan type data “.webm” dikonversi menjadi “.wav” dengan spesifikasi saluran audio *mono channel, sample rate* (laju sampel) 16kHz serta format *encoding* PCM 16-bit. Selanjutnya dilakukan proses secara paralel (*multithread*). Proses paralel pertama (*Thread* 1) dilakukan ekstraksi fitur (ECAPA-TDNN) yang mengubah audio pengguna menjadi vektor *embedding* 192-dimensi. Fitur *Channel Attention* bekerja untuk memilah informasi penting dari *noise* dan menghasilkan biometrik suara pengguna. Selanjutnya dilakukan pencocokan pola (*Pattern Matching*) untuk menghitung jarak *Cosine Similarity* antara biometrik suara pengguna dengan biometrik referensi suara pemilik rumah. Formula *Cosine Similarity* [16] adalah sebagai berikut.

$$CosineSimilarity(A,B) = \frac{A.B}{||A||_2||B||_2} \quad (1)$$

Dimana A adalah biometrik suara pengguna (pemilik rumah dan orang asing) dan B adalah biometrik referensi suara pemilik rumah. Proses paralel kedua (*Thread* 2) mengubah audio pengguna menjadi teks bahasa indonesia dengan menggunakan Google *Speech Recognizer* dan mencari kata kunci "buka" atau "tutup" pagar pada teks tersebut. Jika skor *similarity* lebih besar dari nilai *threshold* dan terdapat kata kunci "buka" atau "tutup" pagar pada teks, maka akses diterima. *Output* berupa keputusan kontrol, sinyal logis (1/0) yang menggerakkan *Motor Stepper* untuk membuka/menutup pagar rumah.

*C. Evaluasi Prototipe*

Prototipe yang selesai dibangun dievaluasi melalui dua pendekatan pengujian.

- Pengujian Kinerja Biometrik Suara: Mengukur kinerja *speaker verification* menggunakan indikator *False Acceptance Rate* (FAR), *False Rejection Rate* (FRR), dan *Equal Error Rate* (EER) [27]. FAR merupakan persentase orang asing yang berhasil membuka/menutup pagar. Semakin rendah nilai FAR, semakin aman sistem dari pembobolan. FRR merupakan persentase pemilik rumah yang gagal membuka/menutup pagar. Nilai FRR yang tinggi akan mengurangi kenyamanan pengguna. Sedangkan EER merupakan persentase keseimbangan antara keamanan dan kenyamanan. Model ECAPA-TDNN dapat dikatakan memiliki kinerja dan sensitivitas yang sangat baik apabila nilai EER yang dihasilkan serendah mungkin. Mengukur FAR, FRR dan ERR akan membuktikan seberapa aman dan seberapa nyaman aplikasi *Smart Voice Gate System* digunakan oleh pengguna.

$$False\ Acceptance\ Rate\ (\text{FAR}) = \frac{FP}{FP+TN}\ x\ 100\% \qquad (2)$$

$$False\ Rejection\ Rate\ (\text{FRR}) = \frac{FN}{FN+TP}\ x\ 100\% \qquad (3)$$

$$Equal\ Error\ Rate\ (\text{EER}) = \frac{FAR+FRR}{2} \qquad (4)$$

Dimana TP (*True Positive*) adalah pemilik rumah yang berhasil membuka/menutup pagar. TN (*True Negative*) adalah orang asing yang gagal membuka/menutup pagar. FP (*False Positive*) adalah orang asing yang berhasil membuka/menutup pagar. FN (*False Negative*) adalah pemilik rumah yang gagal membuka/menutup pagar.

- Pengujian Fungsional (*Black Box Testing*): Memastikan seluruh jalur komunikasi data (Cloudflare *Tunnel*, MQTT, RTSP/LL-HLS) dan fungsi mekanis pagar dapat berjalan.

*D. Perbaikan Prototipe*

Berdasarkan hasil evaluasi awal, perbaikan dilakukan pada pengaturan *threshold* (ambang batas) *Cosine Similarity* untuk mendapatkan nilai EER terendah. Penyesuaian ini memastikan sistem tidak mudah dibobol oleh suara orang asing tanpa mengorbankan kenyamanan pemilik rumah.

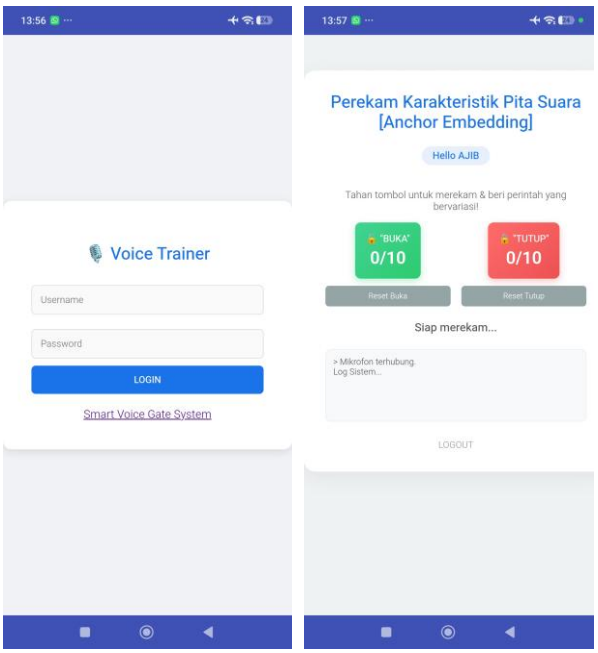
*E. Implementasi Sistem Akhir*

Sistem yang telah direvisi kemudian diimplementasikan secara penuh pada miniatur pagar rumah berbasis IoT *Edge Computing*. Seluruh proses inferensi AI berjalan secara lokal pada Raspberry Pi secara *real-time*.

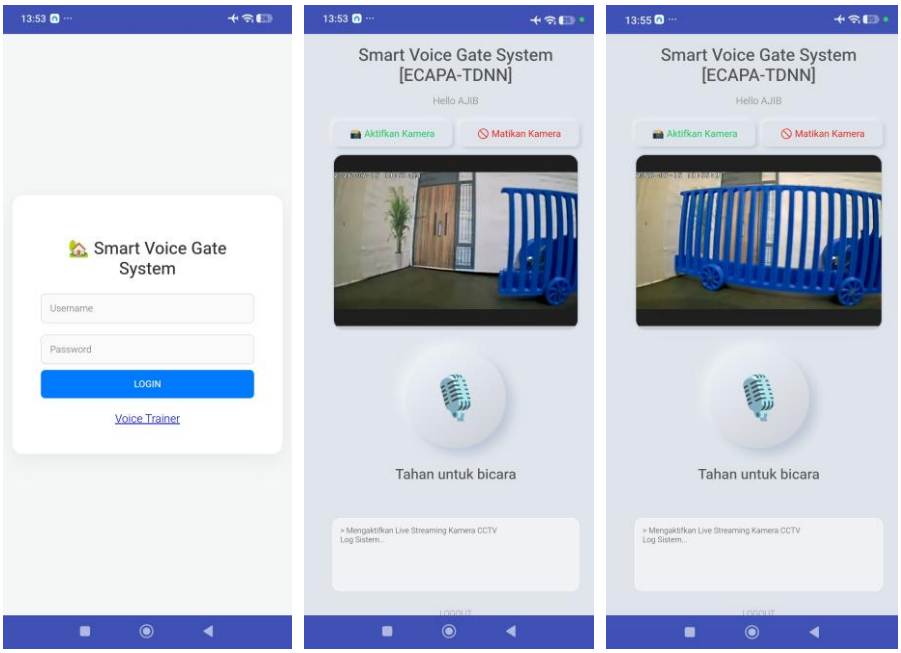
## III. HASIL DAN PEMBAHASAN

*Smart Voice Gate System* menggunakan pendekatan *speaker verification*. Data yang digunakan dibagi menjadi dua kategori, yaitu *Data Enrollment* (Pendaftaran) dan *Data Testing* (Pengujian). *Data Enrollment* diperoleh melalui aplikasi *Voice Trainer.* Aplikasi *Voice Trainer* berfungsi sebagai subsistem pendaftaran biometrik yang mengumpulkan sampel suara pemilik rumah, melakukan prapemrosesan audio, mengekstraksi vektor *embedding* menggunakan model *Deep Learning* ECAPA-TDNN dan menghasilkan *file* biometrik referensi suara pemilik rumah (*anchor embedding*) dengan *type data* ".pt". Sistem merekam suara 1 orang laki-laki pemilik rumah sebanyak 10 rekaman audio buka pagar dan 10 rekaman audio tutup pagar. Sampel ini diproses oleh ECAPA-TDNN untuk menghasilkan satu vektor *embedding* referensi yang disimpan sebagai "Kunci Digital" untuk tiap kata kunci "buka" atau "tutup" pagar. Data yang kedua adalah *Data Testing* (Pengujian), yaitu suara pemilik rumah untuk menguji keberhasilan akses dan suara orang asing untuk menguji keamanan sistem terhadap penyusup. *Data Testing* (Pengujian) diperoleh melalui aplikasi *Smart Voice Gate System.* Aplikasi *Smart Voice Gate System* berbasis android dibangun untuk

mempermudah pengguna dalam mengontrol pagar rumah dan untuk menampilkan pagar rumah dalam kondisi terbuka/tertutup. Aplikasi *voice trainer* dan aplikasi *Smart Voice Gate System* disajikan seperti pada Gambar 5 dan Gambar 6 di bawah ini.




Gambar 5. Aplikasi *Voice Trainer*




Gambar 6. Aplikasi *Smart Voice Gate System*

Pada penelitian ini dilakukan dua jenis pengujian untuk mengukur keberhasilan penelitian, yaitu pengujian kinerja biometrik suara dan pengujian fungsional *(black box testing)*. Pengujian kinerja biometrik suara ECAPA-TDNN yang diimplementasikan pada *Smart Voice Gate System* diukur dengan menggunakan *False Acceptance Rate* (FAR), *False Rejection Rate* (FRR) dan *Equal Error Rate* (EER). Pengujian ini menggunakan suara 1 orang laki-laki pemilik rumah untuk menguji keberhasilan akses dan suara 10 orang asing (5 orang laki-laki dan 5 orang perempuan) untuk menguji keamanan sistem terhadap penyusup. Pengujian dilakukan menggunakan 6 variasi kalimat perintah ("Tolong buka pagar", "Tolong tutup pagar", "Pagar tolong dibuka", "Pagar tolong ditutup", "Buka pagar sekarang juga", "Tutup pagar sekarang juga"). Untuk mengukur FAR, 10 orang asing masing-masing mengucapkan keenam perintah suara tersebut

sebanyak 4x sehingga jumlahnya ada 240 pengujian. Dari 240 pengujian menunjukkan bahwa 9 kali orang asing berhasil membuka/menutup pagar sehingga nilai FAR adalah 3,75%. Sedangkan untuk mengukur FRR, 1 orang laki-laki pemilik rumah mengucapkan keenam perintah suara sebanyak 8x sehingga jumlahnya ada 48 pengujian. Dari 48 pengujian menunjukkan bahwa 1 kali pemilik rumah gagal membuka/menutup pagar sehingga nilai FRR adalah 2,08%. Berdasarkan nilai FAR dan FRR tersebut, nilai EER adalah 2,92%. Hasil pengujian biometrik suara disajikan pada Tabel 1 di bawah ini.

Tabel 1 Hasil Pengujian Biometrik Suara

| Parameter Pengujian | Jumlah Pengujian | Hasil Kinerja Biometrik |
|---|---|---|
| *False Acceptance Rate* (FAR) (10 Orang Asing x 6 Perintah x 4) | 240 Kali | 3,75% (9 Lolos) |
| *False Rejection Rate* (FRR) (1 Pemilik Rumah x 6 Perintah x 8) | 48 Kali | 2,08% (1 Ditolak) |
| *Equal Error Rate* (EER) | Keseimbangan | 2,92% |

Pada pengujian fungsional (*black box testing*) didapatkan hasil sebagai berikut. Audio berhasil terkirim dari HP ke Raspberry Pi melalui Cloudflare *tunnel*. Raspberry Pi berhasil mengirim pesan ke ESP32 melalui protokol MQTT. *Motor stepper* berhasil berputar untuk membuka/menutup pagar saat ESP32 menerima perintah. Kamera CCTV dapat diaktifkan untuk melihat kondisi pagar rumah melalui Cloudflare *tunnel*. Hasil pengujian fungsional (*black box testing*) disajikan pada Tabel 2 di bawah ini.

Tabel 2 Hasil Pengujian Fungsional (*Black Box Testing*)

| Fungsi yang Diuji | Ekspektasi Sistem | Hasil |
|---|---|---|
| Pengiriman suara audio via Cloudflare *Tunnel* | Audio terkirim dari Android ke Raspberry Pi via Cloudflare *Tunnel* | Berhasil |
| Komunikasi MQTT | Raspberry Pi mengirim pesan ke ESP32 melalui protokol MQTT secara *real-time* | Berhasil |
| Aktuasi *Motor Stepper* | Motor berputar membuka/menutup pagar saat verifikasi valid | Berhasil |
| Monitoring CCTV *Live Streaming* | Video kamera CCTV *Live Streaming* lancar via protokol LL-HLS pada aplikasi | Berhasil |

Berdasarkan hasil pengujian kinerja biometrik suara dan pengujian fungsional, integrasi algoritma *deep learning* ECAPA-TDNN pada arsitektur *edge computing* Raspberry Pi terbukti mampu menghadirkan sistem kendali pagar rumah yang berpresisi tinggi dan responsif. Capaian nilai FAR sebesar 3,75%, FRR sebesar 2,08%, dan EER sebesar 2,92% menegaskan keunggulan sistem dalam menyeimbangkan antara keamanan dan kenyamanan bagi pemilik rumah. Didukung dengan keberhasilan pengujian fungsional yang meliputi transmisi sinyal audio dari aplikasi Android melalui Cloudflare *tunnel*, komunikasi MQTT dari Raspberry Pi ke ESP32, eksekusi pergerakan *motor stepper* untuk membuka/menutup pagar, hingga pemantauan *real-time* kondisi pagar lewat CCTV berbasis protokol LL-HLS yang secara efektif mampu mengatasi celah keamanan pada arsitektur *smart home* konvensional yang hanya bergantung pada *speech recognition* tanpa autentikasi biometrik suara.

## IV. KESIMPULAN DAN SARAN

### *A. Kesimpulan*

Berdasarkan hasil perancangan, implementasi, dan pengujian yang telah dilakukan pada *Smart Voice Gate System*, dapat ditarik beberapa kesimpulan sebagai berikut. Penelitian ini berhasil mengintegrasikan algoritma *High-End* ECAPA-TDNN ke dalam perangkat *embedded* Raspberry Pi sebagai *inference engine* pemrosesan biometrik suara secara lokal, yang terhubung secara penuh dengan mikrokontroler ESP32 dan aktuator motor *stepper* via protokol MQTT. Pengujian biometrik suara yang melibatkan 1 orang laki-laki pemilik rumah dan 10 orang asing dengan 6 variasi kalimat perintah menghasilkan nilai FAR sebesar 3,75%, FRR sebesar 2,08%, dan EER sebesar 2,92%. Nilai ini membuktikan bahwa sistem akurat dalam mengenali pemilik rumah sekaligus tangguh dalam menolak penyusup. Penerapan pendekatan verifikasi ganda, yang menggabungkan *speech recognition* dan *speaker verification* berhasil menutupi kelemahan sistem konvensional berbasis *cloud* (seperti *Google Assistant*). Seluruh fitur fungsional yang meliputi pengiriman suara audio via Cloudflare *tunnel*, komunikasi MQTT, aktuasi motor *stepper* dan monitoring CCTV *live streaming* berlatensi rendah berjalan secara stabil. Pemrosesan inferensi AI secara lokal (*Edge AI*) berhasil mengurangi ketergantungan pada *cloud* pihak ketiga, sehingga meningkatkan privasi data pengguna dan responsivitas kendali aktuator secara *real-time*.

*B. Saran*

Untuk pengembangan dan penyempurnaan penelitian *Smart Voice Gate System* di masa mendatang, terdapat saran yang dapat diterapkan. Perlu dilakukan pengujian lebih lanjut dengan mendaftarkan sampel suara dari banyak pengguna sekaligus dalam satu rumah seperti anggota keluarga dengan *gender* dan kelompok usia yang bervariasi untuk menguji keandalan sistem pada skala multi-*user*.

## V. DAFTAR PUSTAKA


[1] Abdurrozaq, G. I., Dewangga, V. D., Ekapagliuca, E. P., Yudhistira, A. M., & Susanto, R. (2024). Alat kunci pintu otomatis berbasis suara dan aplikasi. *Seminar Nasional Teknologi Informasi dan Bisnis (SENATIB)*, 269–275.

[2] Amin, D. H. M. (2022). Voice controlled home automation system. *International Journal of Electrical, Electronics and Computers*, *7*(5). https://dx.doi.org/10.22161/eec.75.1

[3] Ariff, M. I. M., Fadzir, F. D. M., & Salleh, K. A. (2022). Design and development of a smart garage door. *2022 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*, 1–6. https://doi.org/10.1109/IEMTRONICS55184.2022.9795768

[4] Bayerl, S. P., Wagner, D., Baumann, I., Bocklet, T., & Riedhammer, K. (2023). Detecting vocal fatigue with neural embeddings. *Journal of Voice, 39*(4). https://doi.org/10.1016/j.jvoice.2023.01.012

[5] Bhaganagare, S., Chavan, S., Gavali, S., & Godase, V. V. (2025). Voice-controlled home automation with ESP32: A systematic review of IoT-based solutions. *Journal of Microprocessor and Microcontroller Research*, *2*(3), 1–13.

[6] Brydinskyi, V., Khoma, Y., Sabodashko, D., Podpora, M., Khoma, V., Konovalov, A., & Kostiak, M. (2024). Comparison of modern deep learning models for speaker verification. *Applied Sciences*, *14*(4), 1329. https://doi.org/10.3390/app14041329

[7] C. A. Angger Gindaong, I. Nirmala, and C. Suhery. (2026). Application Of Kalman Filter Algorithm to Reduce Noise In Door Lock Opening System With Voice Recognition. *SMARTICS*, *12*(1), 85–95.

[8] Dawalatabad, N., Ravanelli, M., Grondin, F., Thienpondt, J., Desplanques, B., & Na, H. (2021). ECAPA-TDNN embeddings for speaker diarization. *arXiv preprint arXiv:2104.01466*.

[9] Desplanques, B., Thienpondt, J., & Demuynck, K. (2020). ECAPA-TDNN: Emphasized Channel Attention, Propagation and Aggregation in Time Delay Neural Network based speaker verification. *Interspeech 2020*, 3830–3834. https://doi.org/10.21437/Interspeech.2020-2650

[10] He, Q., Zeng, Q., & Zhao, X. (2024). The speaker verification system based on TDNN improvements in noisy environments. *Electronic Engineering and Informatics*, 135–141. https://doi.org/10.3233/ATDE240068

[11] Ihsan, I. P., Buwarda, S., Novianty, H., & Putra, I. A. (2021). Voice recognition untuk otomatisasi sistem pengakses pintu. *JSAI: Journal Scientific and Applied Informatics*, *4*(1), 116–125.

[12] Irugalbandara, C., Naseem, A. S., Perera, S., Kiruthikan, S., & Logeeshan, V. (2023). A secure and smart home automation system with speech recognition and power measurement capabilities. *Sensors*, *23*(13), 5784. https://doi.org/10.3390/s23135784

[13] Karim, D. A., & Anwar, N. (2023). Smart home berbasis IoT menggunakan Arduino Uno dan suara pada Google Assistant. *IKRAITH-INFORMATIKA*, *7*(3). https://doi.org/10.37817/ikraith-informatika.v7i3

[14] Lantz, K.E. (1986). The Prototyping Methodology. *Prentice-Hall*.

[15] Lenni, & Rifa'i, M. A. R. (2022). Prototipe sistem kontrol suara dan monitoring pada pintu pagar berbasis IoT menggunakan NodeMCU ESP8266. *Jurnal Teknik Elektro*, *6*(1). https://dx.doi.org/10.31000/jte.v6i1.6960

[16] Moura, A. A., Nepomuceno, N., & Furtado, V. (2024). Enhancing speaker identification in criminal investigations through clusterization and rank-based scoring. *Forensic Science International: Digital Investigation*, *49*, 301765. https://doi.org/10.1016/j.fsidi.2024.301765

[17] Netinant, P., Utsanok, T., Rukhiran, M., & Klongdee, S. (2024). Development and assessment of Internet of Things-driven smart home security and automation with voice commands. *IoT*, *5*(1), 79–99. https://doi.org/10.3390/iot5010005

[18] Nuqsyahbandi, Pasaribu, F. I., Kusuma, B. S., Evalina, N., Nasution, E. S., Rizky, F., & Sipahutar, E. (2024). Desain sistem kontrol jarak jauh pintu pagar melalui Internet of Things. *Sudo Jurnal Teknik Informatika*, *3*(2). https://doi.org/10.56211/sudo.v3i2.507

[19] Plaza R, M. A. J., & Hanum Maghfiro Risky Ningtias. (2023). Penerapan internet of things pada prototype smart home menggunakan pola suara dengan mikrokontroler nodemcu. *Sienna, 4*(2), 87–96. https://doi.org/10.47637/sienna.v4i2.981

[20] Ravanelli, M., Parcollet, T., Plantinga, P., Rouhe, A., Cornell, S., Lugosch, L., & Bengio, Y. (2021). SpeechBrain: A general-purpose speech toolkit. *arXiv preprint arXiv:2106.04624*.
[21] Samuel, F. A., Modupe, A. O., Titilayo, A. O., Oyeladun, M. B., Abiodun, A. O., Mayowa, I. R., & Samuel, A. M. (2021). Voice recognition system for door access control using mobile phone. *International Journal of Science and Engineering Applications*, *10*(9), 132–139.
[22] Shen, L., Xiong, Y., Guo, D., Mo, W., Yu, L., Yang, H., & Lin, Y. (2026). Attention-based multi-level feature fusion for voice disorder diagnosis. *Measurement*, *258*, 119168.
[23] Sigona, F., & Grimaldi, M. (2024). Validation of an ECAPA-TDNN system for forensic automatic speaker recognition under case work conditions. *Speech Communication*, *158*, 103045. https://doi.org/10.1016/j.specom.2024.103045
[24] Syaputra, A., Samsumar, L. D., & Efendi, M. M. (2024). Rancang bangun sistem akses pintu gerbang berbasis IoT dengan perintah suara. *Journal of Computer Science and Information Technology (JCSIT)*, *2*(1), 61–70.
[25] Thienpondt, J., Desplanques, B., & Demuynck, K. (2020). The IDLAB VoxCeleb speaker recognition challenge 2020 system description. *arXiv preprint arXiv:2010.12468*.
[26] Torad, M. A., Bouallegue, B., & Ahmed, A. M. (2022). A voice controlled smart home automation system using artificial intelligent and internet of things. *TELKOMNIKA Telecommunication Computing Electronics and Control*, *20*(4), 808–816. https://doi.org/10.12928/TELKOMNIKA.v20i4.23763
[27] Zaidi, A. Z., Chong, C. Y., Jin, Z., Parthiban, R., & Sadiq, A. S. (2021). Touch-based continuous mobile device authentication: State-of-the-art, challenges and opportunities. *Journal of Network and Computer Applications, 191*, 103162. https://doi.org/10.1016/j.jnca.2021.103162