

EdgeClap: Implementasi Sistem Otentikasi Pintu Pintar Berbasis Tepuk Tangan Menggunakan Model LSTM yang Dioptimalkan dengan Knowledge Distillation pada Perangkat IoT

Prabu Aditya Atmaja

Fakultas Informatika

Universitas Telkom

Bandung

prabuaditya@students.telkomuniversity.ac.id

Aji Gautama Putrada

Fakultas Informatika

Universitas Telkom

Bandung

ajigps@telkomuniversity.ac.id

Hassan Rizky Putra Sailallah

Fakultas Informatika

Universitas Telkom

Bandung

hassanrizkyhrs@telkomuniversity.ac.id

Apa. Penelitian ini mengembangkan dan mengevaluasi sistem otentikasi pintu pintar berbasis pengenalan pola tepuk tangan menggunakan model LSTM yang dioptimalkan dengan metode Knowledge Distillation (KD) agar dapat diimplementasikan pada perangkat IoT seperti ESP32.

Mengapa. Sistem smart lock saat ini pada umumnya masih banyak yang menggunakan dan bergantung pada kamera, sensor optik, atau perangkat wearable yang mahal. Sementara sistem berbasis sensor suara saat ini masih banyak yang menggunakan metode threshold sederhana dan belum dapat membedakan suara pola tepuk tangan secara pintar. Oleh karena itu diperlukan pendekatan yang ringan, cerdas, dan dapat berjalan langsung tanpa koneksi internet, serta dengan biaya yang relatif lebih murah.

Bagaimana. Pengumpulan dataset dilakukan dengan menjalankan sensor KY-037 untuk mendapatkan data dari dua suara berbeda, yaitu suara tepuk tangan dan non-tepuk tangan. Kemudian dilakukan pelatihan model LSTM berdasarkan dataset yang sudah didapatkan, model utama ini akan menjadi teacher model yang nantinya dioptimasi menggunakan metode KD untuk menghasilkan model student yang lebih ringan. Perbandingan performa dilakukan menggunakan confusion matrix, jumlah parameter, ukuran model, dan inference time. Model student kemudian diimplementasikan pada perangkat IoT seperti ESP32 untuk inferensi real-time

Hasil. Student model mendapatkan akurasi validation accuracy sebesar 0.9731, nilai ini lebih tinggi dari teacher model yang hanya mendapatkan nilai sebesar 0.9676. Metode KD juga berhasil mereduksi ukuran model sebesar 57.1% dan jumlah parameter sebesar 25.8%, serta mampu melakukan inferensi realtime sekitar 68ms di atas perangkat IoT tanpa menggunakan internet maupun cloud computing.

Kata kunci: otentikasi akustik, smart lock, LSTM, knowledge distillation, IoT, edge computing.

Several studies have shown that biometric-based authentication offers advantages and increased security. However, the performance of clap-based authentication with voice sensors can be improved by artificial intelligence (AI) methods such as long short-term memory (LSTM). This study presents EdgeClap, a smart door lock authentication system based on LSTM-based voice-clap pattern recognition. First, we designed the architecture of the voice-based smart lock system and implemented it. Then, we collected a dataset with several voice labels and used it to train the LSTM. This model was then trained and optimized using the Knowledge Distillation (KD) technique. We then compared the prediction performance and model size between LSTM and LSTM + KD using a confusion matrix, number of parameters, model size, and inference time. The test results showed that the KD method provides efficiency while maintaining prediction performance. The student model achieved a validation accuracy of 0.9731, higher than the teacher model's 0.9676. In addition, KD improved efficiency by reducing the model size by 57.1% and the number of parameters by 25.8%. This optimization enables deep learning models to be implemented directly on the ESP32, enabling real-time inference 68ms without reliance on the internet, cloud computing, or wearable devices.

Keywords: acoustic authentication, smart lock, long short-term memory, knowledge distillation, Internet of Things, edge computing

1. Pendahuluan

Latar Belakang

Perkembangan teknologi khususnya IoT sangat berkembang pesat akhir-akhir ini, IoT telah membawa manfaat dan perubahan yang cukup signifikan dalam kehidupan sehari-hari [1], contoh dari penggunaan IoT

dalam kehidupan sehari-hari ini adalah sistem otomatisasi rumah dan sistem keamanannya yang berupa *smart lock door*. Sistem penguncian pintu pintar ini menawarkan beberapa cara untuk dalam pengaplikasian nya, beberapa pilihan di antaranya yaitu dengan menggunakan RFID, otentikasi PIN, sidik jari, hingga pengenalan wajah [2]. Akan tetapi sebagian besar dari hal tersebut memerlukan interaksi fisik ataupun perangkat tambahan seperti kamera, dan sensor optik. Hal ini mendorong adanya inovasi dalam bidang IoT untuk dapat mengembangkan dan mengintegrasikan IoT agar lebih alami, seperti dengan melakukan pemrosesan yang ringan dan dilakukan langsung pada *IoT edge device* tanpa bergantung pada *server cloud*.

Beberapa penelitian sebelumnya sudah menunjukkan potensi yang lumayan besar dari penggunaan sensor suara dalam sistem kendali IoT. Dalam penelitian oleh Kesadaran Luaha *et al.* [3] dengan judul “Design of a Clap Detection Light Control System Based on Arduino Uno” berhasil mengimplementasikan sensor suara FC-04 untuk mendeteksi tepuk tangan yang digunakan untuk menyalakan atau mematikan lampu. Namun, sistem tersebut masih menerapkan metode berbasis *threshold* atau ambang batas yang membuatnya belum dapat membedakan pola tepuk tangan dengan kebisingan lingkungan dengan cerdas. Di sisi lain, penelitian oleh Qonita Banafsa [4] dengan judul “Implementasi Sistem Smart Door Lock Berbasis Pengenalan Wajah dengan Kamera Huskylens” menunjukkan pendekatan yang berbeda dan menggunakan Artificial Intelligence AI dalam penelitian ini, yaitu sistem menggunakan pengenalan wajah berbasis AI dalam membuka pintu sebagai otentikasi tambahan. Meski sistem ini berhasil meningkatkan keamanan secara signifikan, namun biaya dalam pembuatan dan pengembangan alat tergolong cukup mahal, dan masih memiliki beberapa kelemahan dalam sistemnya seperti jika dihadapkan dengan situasi pencahayaan yang rendah.

Oleh karena itu, penelitian lebih lanjut yang mengintegrasikan penggunaan *deep learning* seperti model *Long Short-Term Memory* (LSTM) diperlukan agar sistem dapat mengenali pola dari suara tepuk tangan secara temporal dan efisien [5]. Model ini kemudian akan dioptimalkan menggunakan teknik *Knowledge Distillation* (KD) agar menjadi lebih ringan dan dapat dijalankan langsung di atas perangkat IoT[6].

Topik dan Batasannya

Fokus utama pada penelitian ini adalah untuk melakukan pengembangan dan evaluasi terhadap sistem otentikasi *smart lock door* berbasis pengenalan suara tepuk tangan dengan menggunakan pendekatan *deep learning*. Permasalahan yang dikaji adalah bagaimana cara untuk menerapkan model LSTM untuk mengenali pola temporal dari suara tepuk tangan, dan bagaimana mengoptimalkan model tersebut menggunakan teknik KD agar dapat dijalankan pada perangkat IoT. Batasan pada penelitian ini hanya dilakukan pada pengenalan pola tepuk tangan tunggal (*single user*), pengujian yang dilakukan pada area *indoor*, dan sistem yang hanya diuji pada perangkat ESP32 tanpa melibatkan komputasi awan, implementasi difokuskan pada akses kontrol

pintu, dan model dibatasi pada penggunaan arsitektur LSTM.

Tujuan

Tujuan dari penelitian ini adalah untuk mengembangkan dan mengimplementasikan sistem otentikasi *smart door lock* berbasis pengenalan suara yang efisien dan terjangkau, secara spesifik tujuan dari penelitian ini adalah sebagai berikut:

1. Menerapkan model LSTM untuk mengenali dan mengklasifikasi pola temporal pada suara tepuk tangan.
2. Mengevaluasi hasil optimalisasi model LSTM dengan menggunakan teknik KD untuk menghasilkan model yang ringan dan dapat dijalankan pada perangkat IoT.
3. Merancang integrasi antara sistem pengenalan suara tepuk tangan dengan modul pengunci pintu secara lokal dan *real-time*.

Dengan tercapainya tujuan tersebut, penelitian ini diharapkan menghasilkan sistem otentikasi berbasis audio biometrik yang ringan, akurat, dan efisien. Serta memberikan kontribusi terhadap pengembangan teknologi Edge AI pada sistem IoT.

2. Studi Terkait

Beberapa penelitian sebelumnya telah dilakukan eksplorasi dalam penggunaan LSTM, KD, maupun sistem pengenalan suara untuk sistem cerdas. Ringkasan kesenjangan atau *gap* penelitian yang didapati dari beberapa penelitian tersebut antara lain ditunjukkan pada Tabel 1.

Tabel 1. Perbandingan Penelitian Terkait pada Sistem Otentikasi dan Pengenalan Suara

CITE	Pengenalan Tepuk Tangan	LS TM	Knowl edge Distillation	Edge Compu ting	Sm art Loc k
[7]	X	X	✓	✓	X
[8]	X	X	✓	✓	X
[9]	X	✓	X	X	X
[10]	X	✓	X	X	X
[11]	✓	X	X	X	X
Metode yang diusulkan	✓	✓	✓	✓	✓

Beberapa penelitian sebelumnya telah dilakukan dengan menggunakan KD. Pada penelitian yang dilakukan oleh Putrada *et al.* [7] menunjukkan penggunaan KD yang sangat efektif dalam mengompresi model ML, dengan tingkat akurasi terbaik yang didapatkan mencapai 0,93 pada ukuran model target 45 kB. Kemudian pada penelitian lainnya oleh Wang *et al.* [8] mendapatkan hasil yang juga efisien dalam penggunaan KD, pada

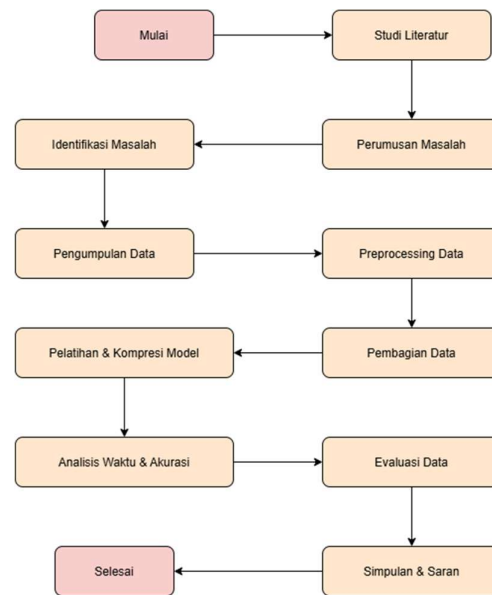
penelitian tersebut menunjukkan penurunan penggunaan memori hingga 96.58% dibandingkan saat menggunakan model *teacher*, akurasi yang tinggi dan ukuran model yang kecil ini sangat cocok untuk disematkan pada lingkungan *edge computing* dengan sumber daya terbatas agar respons terjadi secara *real-time*. Penerapan sistem penguncian pintu berbasis pengenalan suara juga mulai diteliti sebagai alternatif yang lebih interaktif [12].

Di sisi lain, penelitian yang dilakukan oleh Moy'awiah *et al.* [9] dan Ruvina *et al.* [10] menunjukkan keunggulan model LSTM dalam mengklasifikasikan pola suara sekuensial yang kompleks. Kemampuan model tersebut dalam mengingat pola sekuensial dan memori spasial menjadikannya sangat andal dalam mengklasifikasi data suara yang rumit [15]. Meskipun terdapat skema otentikasi menggunakan gestur tepuk tangan yang memerlukan sensor gerak seperti yang diusulkan pada penelitian [11] berupa 'ClapAuth', penggunaan sensor suara juga dapat digunakan untuk melakukan otentikasi [13] dan terbukti jauh lebih murah serta efisien dibandingkan dengan menggunakan sensor gerak yang akan menambah biaya dalam penggunaan perangkat tambahan seperti *smartwatch* yang perlu digunakan pada saat melakukan otentikasi (*wearable device*) [14].

Berdasarkan Tabel 1, dapat terlihat bahwa keseluruhan metode yang diusulkan pada penelitian ini belum digunakan secara keseluruhan juga pada penelitian-penelitian sebelumnya, maka dari itu diharapkan hasil dari penelitian ini dapat membawa kontribusi yang nyata dan bermanfaat terhadap perkembangan ilmu pengetahuan dalam bidang IoT, khususnya pada bidang yang diusulkan dalam penelitian ini seperti penggunaan teknik KD dalam optimalisasi model LSTM yang disematkan pada perangkat IoT.

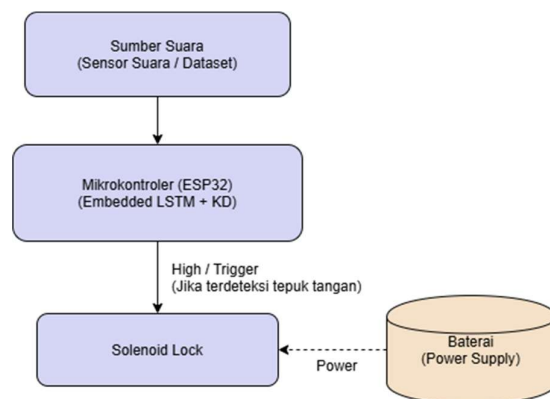
3. Sistem yang Dibangun

Penelitian ini dikembangkan dengan alur yang terstruktur mulai dari Studi Literatur, Pengumpulan Data, Prapemrosesan, pelatihan dan kompresi model, pembagian data, analisis, hingga evaluasi sistem. Data dikumpulkan secara mandiri menggunakan sensor KY-037 dengan merekam suara kondisi ruangan normal dan suara kondisi ruangan dengan tepuk tangan. Alur penelitian secara keseluruhan ditunjukkan oleh Gambar 1.



Gambar 1 Tahapan Penelitian.

Sistem ini dikembangkan dengan menggunakan sensor KY-037 untuk mengumpulkan dataset secara mandiri. Sensor ini dipilih karena kemampuannya mendeteksi suara berdasarkan prinsip elektromagnetik yang sensitif terhadap kenaikan desibel. Setelah tahap prapemrosesan, model LSTM akan dilatih untuk mendapat hasil yang paling optimal, kemudian model tersebut dikompres menggunakan metode KD agar lebih ringan. Arsitektur IoT yang dirancang dapat dilihat pada Gambar 2, mengintegrasikan sensor suara dengan mikrokontroler dan *Solenoid Lock* yang ditenagai oleh Baterai (*Power Supply*), penggunaan baterai terpisah untuk *Solenoid Lock* diperlukan agar menjaga kestabilan daya mikrokontroler dari lonjakan arus.



Gambar 2 Arsitektur Sistem Smart Lock Berbasis Suara

Sistem menerima suara dari sensor atau dataset yang kemudian diteruskan ke mikrokontroler. Kemudian mikrokontroler menerima input berupa suara, yang selanjutnya akan dilakukan klasifikasi apakah suara yang diterima merupakan suara tepuk tangan atau hanya suara ruangan biasa. Setelah pengklasifikasian selesai dan mendapat output, mikrokontroler akan mengirim sinyal *High/Trigger* jika berhasil mengklasifikasi suara tepuk tangan, selanjutnya *solenoid lock* akan menunggu

sinyal dari mikrokontroler, modul ini juga ditenagai oleh Baterai (*Power Supply*) yang berbeda dari mikrokontroler.

Model LSTM dirancang untuk mengatasi masalah *Vanishing Gradient* pada RNN tradisional [16], Kunci utama dari LSTM terdapat pada bagian *cell state* (C_t) yang bekerja sebagai memori utama, dan tiga gerbang yang mengatur aliran informasi. Proses perhitungan pada *Forget Gate* dirumuskan sebagai berikut :

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

Selanjutnya, tahap *Input Gate* menentukan informasi baru dengan lapisan sigmoid (i_t) dan kandidat vektor baru ($\tilde{C}_t$):

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

Pembaruan *Cell State* (C_t) dikalkulasi dengan rumus :

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

Dan *Output Gate* menentukan keluaran *hidden state* (h_t) sebagai berikut :

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

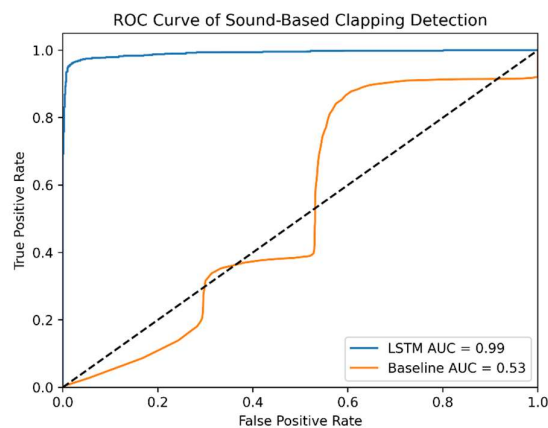
$$h_t = o_t * \tanh(C_t)$$

Untuk pengimplementasiannya di perangkat IoT, model dioptimasi menggunakan metode KD, kemudian model hasil kompresi tersebut dikonversi ke format TensorFlow Lite (.tflite), lalu diubah ke dalam bentuk *C byte array* (.h) agar dapat dijalankan secara lokal di atas ESP32. Metrik kinerja yang diukur berdasarkan *Confusion Matrix*, yang mencakup Akurasi, Presisi, *Recall*, dan *F1 Score*. Metrik efisiensi diukur menggunakan parameter waktu inferensi ($\Delta T = t_{end} - t_{start}$) serta perbandingan ukuran model dari hasil kompresi dan jumlah parameter.

4. Evaluasi

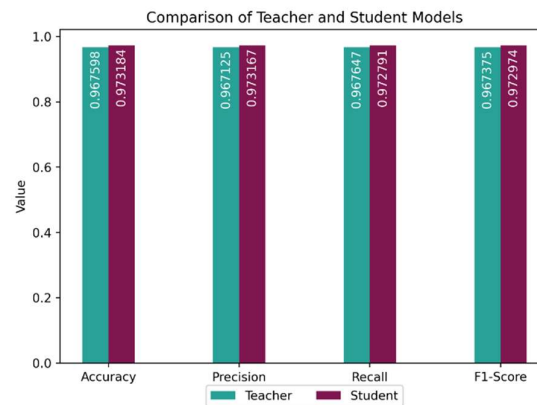
4.1 Hasil Pengujian Skenario 1: Kinerja Klasifikasi

Skenario ini mengevaluasi hasil dari kemampuan model untuk mengklasifikasikan suara tepuk tangan dan non-tepuk tangan. Dataset yang diperoleh secara mandiri mendapatkan jumlah sebanyak 8967 data, dengan 4708 di antaranya berasal dari pengambilan suara ruangan dengan adanya tepuk tangan, dan 4259 lainnya berasal dari hasil pengambilan suara ruangan tanpa adanya tepuk tangan (kondisi ruangan normal). Data dinormalisasi dalam skala [0,1] dan dibagi menjadi 80% data latih dan 20% data uji. Parameter pelatihan menggunakan *time step* sebesar 20.



Gambar 3. Kurva ROC dari pendeteksian akustik tepuk tangan.

Hasil perbandingan nilai ROC memperlihatkan bahwa model LSTM berhasil memperoleh nilai AUC sebesar 0.99, hal ini menunjukkan bahwa model LSTM jauh lebih mampu membedakan suara tepuk tangan dan non-tepuk tangan dibandingkan dengan metode *Baseline* (*threshold*) yang hanya memperoleh nilai AUC sebesar 0.53.

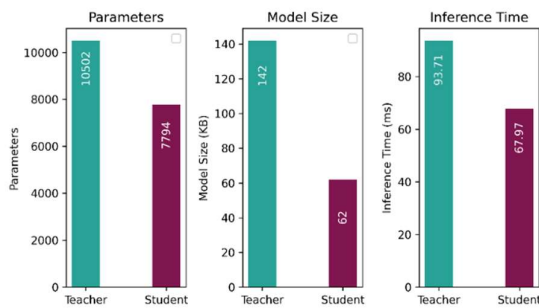


Gambar 4. Komparasi performa antara model guru dengan murid.

Pada Gambar 4, terlihat grafik dari hasil komparasi performa antara model guru dengan murid, setelah dilakukannya pelatihan sebanyak 100 *Epoch* untuk *Teacher Model* dan 30 *Epoch* untuk *Student Model*, komparasi menunjukkan akurasi *Student Model* mencapai angka sekitar 0.9731, mengungguli *Teacher Model* yang hanya mendapatkan akurasi sebesar kurang lebih 0.9676. Nilai *Confusion Matrix* dari *Student Model* juga mencatat 958 *True Positive*, 14 *False Positive*, 806 *True Negative*, dan 12 *False Negative*.

4.2 Hasil Pengujian Skenario 2: Efisiensi Komputasi

Skenario ini memvalidasi efektivitas penggunaan teknik KD, nilai parameter *Temperature* yang digunakan pada penelitian ini adalah 5.0 dan bobot (α) sebesar 0.5. Pengujian dilakukan pada ESP32 dengan *input buffer* sebanyak 20 data agar model dapat dengan mudah melakukan klasifikasi.



Gambar 5. Komparasi efisiensi antara model guru dengan murid.

Gambar 5 menunjukkan grafik hasil komparasi efisiensi antara *Teacher Model* dengan *Student Model*, terlihat dari segi efisiensi komputasi, jumlah parameter dari *Teacher Model* berhasil tereduksi sekitar 25.8%, dari yang tadinya bernilai 10502, menjadi hanya 7794, kemudian ukuran model juga tereduksi dengan sangat drastis sebesar kurang lebih 57.1%, dari 142 KB menjadi hanya sekitar 62 KB. Waktu inferensi yang dimiliki *Student Model* juga lebih kecil dibandingkan dengan *Teacher Model*, *Student Model* mendapat waktu inferensi kurang lebih sebesar 67.97 ms, sedangkan *Teacher Model* memiliki waktu inferensi kurang lebih sebesar 93.71 ms.

4.3 Analisis Hasil Pengujian

Pada skenario 1, penggunaan model LSTM terbukti jauh lebih mampu dalam membedakan antara suara tepuk tangan dengan non-tepuk tangan dikarenakan model LSTM melihat urutan waktu (*Temporal Dependencies*). Sedangkan metode *Baseline* hanya melihat amplitudo dari inputan yang masuk tanpa melihat urutan waktu kebelakang, sehingga lebih mudah tertipu suara lonjakan desibel dan melakukan kesalahan dalam mengambil keputusan apakah suara yang ditangkap merupakan suara tepuk tangan atau hanya kebisingan latar saja. Selanjutnya analisis mengenai tingkat akurasi yang dimiliki *Student Model* yang mengungguli *Teacher Model*, hal ini sejalan dengan teori jika kita dapat mengurangi atau bahkan menghilangkan bias pada dataset awal, akurasi dan hasil dari *Student Model* akan jauh lebih bagus dibanding *Teacher Model* karena *Student Model* tidak hanya menghafal data melainkan juga mempelajari pola abstraksi yang lebih bersih [17]. Kemudian nilai dari *Confusion Matrix*, dimana nilai *False Positive* yang rendah (14) dari *Student Model* memastikan pintu jarang terbuka sendiri tanpa suara tepuk tangan, dan nilai *False Negative* yang minim (12) membuktikan sistem jarang gagal merespons pengguna. Pada skenario 2, reduksi ukuran model dari 142 KB menjadi 62 KB, dan penurunan parameter sebesar 25.8% bukanlah sekedar angka matematis. ESP32 memiliki batasan SRAM yang sangat ketat, reduksi ini memastikan model *Deep Learning* muat dan dapat berjalan langsung di atas ESP32 tanpa memerlukan memori yang banyak, sekaligus menyisakan ruang memori untuk tugas lain seperti menggerakkan *Solenoid Lock*. Selanjutnya waktu inferensi dari *Student Model* yang mendapat angka di bawah 70 ms memastikan sistem 'EdgeClap' mampu beroperasi secara *real-time*. Semua komputasi terjadi secara offline di perangkat *end*

device, sehingga terbebas dari bahaya latensi serta intervensi jaringan luar, dan privasi penghuni rumah juga terjaga.

5. Kesimpulan

Penelitian ini mengusulkan dan mengevaluasi sistem otentikasi pintu pintar berbasis tepuk tangan. Berdasarkan hasil pengujian, model LSTM menunjukkan kinerja yang sangat bagus dalam melakukan klasifikasi terhadap suara tepuk tangan di antara suara kebisingan, model ini memperoleh nilai AUC mencapai 0.99. Penerapan metode *Knowledge Distillation* (KD) juga sukses dalam melakukan kompresi terhadap model dengan penurunan ukuran sebesar 57.1% dan parameter sebesar 25.8%, penurunan ukuran dan parameter ini juga dibarengi dengan penaikan akurasi validasi model hingga mencapai sekitar 0.9731. Integrasi *Edge Computing* juga berhasil diimplementasikan pada ESP32 dan mendapat waktu inferensi yang cukup cepat sekitar 67.97 ms, membuktikan bahwa sistem 'EdgeClap' sangat cocok untuk dijalankan pada perangkat IoT dengan sumber daya yang terbatas.

Untuk pengembangan selanjutnya, disarankan untuk mengembangkan sistem yang dapat menerapkan pola tepuk tangan yang bervariasi, pola tepuk tangan yang spesifik untuk tiap individu, agar tingkat keamanan lebih personal. Diperlukan juga uji ketahanan ke area *outdoor* dengan tingkat kebisingan yang cenderung bervariasi dan naik turun atau ruangan dengan pantulan gema yang tinggi. Terakhir, penelitian lanjutan disarankan untuk dapat melakukan komparasi dengan model *lightweight* lainnya, untuk melihat apakah ada alternatif lain yang dapat menawarkan efisiensi atau akurasi yang lebih tinggi dibanding kombinasi LSTM+KD pada ekosistem ESP32.

REFERENSI

- [1] Y. B. Zikria, R. Ali, M. K. Afzal, and S. W. Kim, "Next-generation internet of things (iot): Opportunities, challenges, and solutions," *Sensors*, vol. 21, no. 4, p. 1174, 2021.
- [2] H. Touqeer, S. Zaman, R. Amin, M. Hussain, F. Al-Turjman, and M. Bilal, "Smart home security: challenges, issues and solutions at different iot layers," *The Journal of Supercomputing*, vol. 77, pp. 14053–14089, 2021.
- [3] K. Luaha, S. Ismail, A. Rahman, N. N. Siagian, L. R. Amalia, and R. F. Wijaya, "Design of a clap detection light control system based on arduino uno," *Zero: Jurnal Sains, Matematika, dan Terapan*, vol. 6, no. 2, pp. 142–146, 2022.
- [4] Q. Banafsa, Implementasi Sistem Smart Door Lock Berbasis Pengenalan Wajah dengan Kamera Huskylens, N. Hendrarini and D. N. Ramadan, Eds. Bandung: Universitas Telkom, D3 Teknologi Komputer, 2025, [Online]. Available: <https://openlibrary.telkomuniversity.ac.id/home/catalog/id/238057/>.
- [5] A. Altayeva, M. Omarov, and H.-N. Lee, "Convolutional lstm network for real-time impulsive sound detection and classification in urban

environments,” *Applied Sciences*, vol. 13, no. 2, p. 721, 2023.

[6] A. Alkhulaifi, F. Alsahli, and I. Ahmad, “Knowledge distillation in deep learning and its applications,” *PeerJ Computer Science*, vol. 7, p. e474, 2021.

[7] A. G. Putrada, M. Abdurrohman, D. Perdana, and H. H. Nuha, “Edgesl: Edge-computing architecture on smart lighting control with distilled knn for optimum processing time,” *IEEE Access*, vol. 11, pp. 64697–64712, 2023.

[8] Y. Wang, Z. Yu, J. Wu, C. Wang, Q. Zhou, and J. Hu, “Adaptive knowledge distillation-based lightweight intelligent fault diagnosis framework in iot edge computing,” *IEEE Internet of Things Journal*, vol. 11, no. 13, pp. 23156–23169, 2024.

[9] M. A. Al-Shannaq, A. Nasrawi, A. A.-R. K. Bsoul, and A. A. Saifan, “Abnormal heart sound recognition using svm and lstm models in real-time mode,” *Scientific Reports*, vol. 15, p. 9129, 2025.

[10] S. Ruvinga, G. J. Hunter, O. Duran, and J.-C. Nebel, “Use of lstm networks to identify “queenlessness” in honeybee hives from audio signals,” in *2021 17th International Conference on Intelligent Environments (IE)*, pp. 1–6, 2021.

[11] A. Buriro and F. Ricci, “Clapauth: A gesture-based user-friendly authentication scheme to access a secure infrastructure,” in *Emerging Technologies for Authorization and Authentication: 5th International Workshop, ETAA 2022, Copenhagen, Denmark,*

September 30, 2022, Revised Selected Papers, (Berlin, Heidelberg), p. 15–30, Springer-Verlag, 2022.

[12] M. Syafrizal, V. Ikawati, A. Siswanto, N. Lestari et al., “Sistem keamanan door lock berbasis voice recognition dengan natural language processing,” *Infotronik: Jurnal Teknologi Informasi Dan Elektronika*, vol. 9, no. 1, pp. 1-11, 2024.

[13] A. Jylh’a, C. Erkut, U. Simsekli, and A. Cemgil, “Sonic handprints: Person identification with hand clapping sounds by a model-based method,” 01 2012.

[14] D. Y. Z. Chan and B. S. Tan, “Pentap: A novel rhythm game with innovative acoustic soundwave-based input and audio recognition,” in *2024 3rd International Conference on Digital Transformation and Applications (ICDXA)*. IEEE, 2024, pp. 45-50.

[15] T. A. Onisha, J. Kim, and J. Seol, “Multi label sound classification using deep learning models,” in *2024 IEEE/ACIS 22nd International Conference on Software Engineering Research, Management and Applications (SERA)*. IEEE, 2024, pp. 129-134.

[16] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, pp. 1735–1780, 11 1997.

[17] J. Zhang, Y. Gao, R. Liu, X. Cheng, H. Zhang, and S. Chen, “Can students beyond the teacher? distilling knowledge from teacher’s bias,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, no. 21, 2025, pp. 22 434–22 442.