

IMPLEMENTASI JARINGAN HIERARKI ATTENTION UNTUK KLASIFIKASI BASIS DATA MULTIMODAL BIOMETRIK

David Bona¹⁾, Betha Nurina Sari²⁾

^{1,2)} Universitas Singaperbangsa Karawang

Jl. HS.Ronggo Waluyo, Puseurjaya, Kec. Telukjambe Tim., Kabupaten Karawang, Jawa Barat, Indonesia

e-mail: davidbonapurba@gmail.com¹⁾, betha.nurina@staff.unsika.ac.id²⁾

ABSTRAK

Pada era digital ini, Teknologi semakin pesat termasuk di bidang Machine Learning, bekerja pada Machine Learning (ML) telah dipercepat dalam beberapa tahun terakhir. Klasifikasi dapat dilakukan dengan banyak metode, salah satunya yaitu menggunakan Deep Learning. Deep Learning merupakan bagian dari Machine Learning yang dapat mempelajari metode komputasinya sendiri. Deep Learning yang saat ini memiliki hasil paling signifikan dalam pengenalan citra adalah Convolutional Neural Network (CNN) salah satunya sistem identifikasi menjadi hal yang penting untuk dapat mengenali seseorang berdasarkan ciri yang dimilikinya. Pada umumnya sistem identifikasi seseorang masih dilakukan menggunakan teknik konvensional, seperti menggunakan ID card yang mudah rusak dan hilang, kata sandi atau PIN yang mudah dilupakan. Saat ini, biometrik banyak digunakan untuk verifikasi dan identifikasi. Keunikan dan ketersediaan data biometrik membuatnya ideal untuk tujuan ini. Informasi biometrik biasanya mencakup fitur wajah, sidik jari, iris mata, tanda tangan, suara, dan DNA. Dalam hal ini erat kaitannya dengan sistem otentikasi biometrik multimodal yang terbukti menjadi pendekatan yang menjanjikan untuk mengidentifikasi seseorang menggunakan fitur fisiologis dan/atau perilaku seperti wajah, sidik jari, geometri tangan, iris, retina, vena, serta ucapan. Biometrik multimodal wajah, telapak tangan, telapak tangan dan audio memiliki beberapa keunggulan termasuk pemanfaatan akuisisi gambar non-invasif dan berbiaya rendah. Selain itu, gambar wajah dan sidik jari dapat dengan mudah diperoleh secara bersamaan menggunakan dua sensor tanpa sentuhan. Pada penelitian ini dalam klasifikasinya menggunakan jaringan Hierarki Attention sehingga memperoleh Hasil dari evaluasi matrik dari implementasi jaringan hierarki attention untuk klasifikasi basis data multimodal biometrik diperoleh akurasi sebesar 100%

Kata Kunci: Deep learning, Biometrik Multimodal, Jaringan Hierarki Attention

ABSTRACT

In this digital era, technology is growing rapidly, including in the field of machine learning, work on machine learning (ML) has accelerated in recent years. Classification can be done by many methods, one of which is using Deep Learning. Deep Learning is part of Machine Learning that can learn its own computational methods. Deep Learning which currently has the most significant results in image recognition is the Convolutional Neural Network (CNN), one of which is an identification system that is important to be able to recognize someone based on their characteristics. In general, a person's identification system is still carried out using conventional techniques, such as using ID cards that are easy to break and lose, passwords or PINs that are easy to forget. Currently, biometrics are widely used for verification and identification. The uniqueness and availability of biometric data makes it ideal for this purpose. Biometric information typically includes facial features, fingerprints, irises, signatures, voice, and DNA. In this case, it is closely related to the multimodal biometric authentication system which is proving to be a promising approach to identify a person using physiological and/or behavioral features such as face, fingerprint, fingerprint, hand geometry, iris, retina, vein, and speech. Multimodal facial, palm, palm and audio biometrics have several advantages including the utilization of low-cost, non-invasive image acquisition. In addition, face and fingerprint images can be easily obtained simultaneously using the two touchless sensors. In this study, the classification uses the Attention Hierarchy network so as to obtain the results from the evaluation of the matrix from the implementation of the attention hierarchy network for the classification of multimodal bio-metric databases, an accuracy of 100% is obtained.

Keywords: Deep learning, Multimodal Biometrics, Attention Hierarchy Network

I. PENDAHULUAN

Perkembangan Teknologi semakin pesat termasuk di bidang *Machine Learning*, bekerja pada *Machine Learning* (ML) telah dipercepat dalam beberapa tahun terakhir. Klasifikasi dapat dilakukan dengan banyak metode, salah satunya yaitu menggunakan *Deep Learning*. *Deep Learning* merupakan bagian dari *Machine Learning* yang dapat mempelajari metode komputasinya sendiri. *Deep Learning* yang saat ini memiliki hasil paling signifikan dalam pengenalan citra adalah *Convolutional Neural Network* (CNN)[1]. Hal tersebut dikarenakan CNN berusaha meniru sistem pengenalan citra pada visual cortex manusia sehingga memiliki kemampuan mengolah informasi citra. Namun, CNN seperti metode *Deep Learning* lainnya, memiliki kelemahan yaitu proses pelatihan model

yang lama. Dengan perkembangan perangkat keras, hal tersebut dapat diatasi menggunakan teknologi *General Purpose Graphical Processing Unit* (GPGPU). CNN dirancang khusus untuk pengenalan dan klasifikasi gambar. CNN memiliki beberapa lapisan (layer) yang mengekstrak informasi dari gambar dan menentukan klasifikasi dari gambar berupa skor klasifikasi[2]. Karena kemampuannya untuk menambang, menganalisis, dan menafsirkan file besar, metode *Deep Learning* telah diterapkan ke banyak disiplin ilmu termasuk pemrosesan gambar dan biometrik. Dasar untuk setiap metode biometrik adalah beberapa jenis metode pencocokan yang biasanya satu lawan satu dalam hal verifikasi dan satu ke banyak dalam hal identifikasi. Di sisi lain, metode Klasifikasi *Deep Learning* telah terbukti meningkatkan kinerja sistem biometrik dan potensi penuhnya dalam membantu biometrik mencapai kinerja 100% belum dieksplorasi secara memadai[3].

Saat ini, biometrik banyak digunakan untuk verifikasi dan identifikasi. Keunikan dan ketersediaan data biometrik membuatnya ideal untuk tujuan ini. Informasi biometrik biasanya mencakup fitur wajah, sidik jari, iris mata, tanda tangan, suara, dan DNA. Aplikasi untuk biometrik termasuk smartphone, yang menggunakan sidik jari atau pengenalan wajah untuk membuka kunci layar dan pengenalan wajah atau sensor sidik jari untuk memverifikasi jam kerja. Meskipun biometrik adalah fitur unik, biometrik dapat mengalami tingkat kesalahan, seperti tingkat kesalahan yang tidak dapat diterima atau penipuan berbahaya. Sistem pengenalan biometrik dianggap sebagai alat yang ampuh untuk mengidentifikasi individu melalui pendekatan semi-otomatis atau sepenuhnya otomatis berdasarkan ciri-ciri perilaku (misalnya suara dan tanda tangan) dan/atau ciri-ciri fisik (misalnya iris mata dan sidik jari). Sifat unik data biometrik dipandang lebih unggul daripada pendekatan pengenalan tradisional (misalnya kata sandi) karena tidak dapat hilang, dicuri, atau direplikasi[4].

Metode otentikasi berdasarkan teknologi biometrik menjamin bahwa alat identifikasi tidak dapat dicuri, hilang atau dilupakan, semakin dituntut dalam lingkungan dan aplikasi keamanan seperti kontrol akses dan transaksi elektronik. Upaya besar telah dilakukan di komunitas biometrik untuk meningkatkan keandalan sistem keamanan dengan menggabungkan yang ada pengenalan ahli unimodal untuk menerapkan sistem otentikasi multimodal yang sesuai dengan kebutuhan pasar[5]. Namun demikian, dalam situasi dunia nyata, tingkat kesalahan yang dicapai dengan teknologi mutakhir telah memperlambat aplikasi umum yang telah dikembangkan sebelumnya. Untuk mengatasi perbedaan kinerja antara eksperimen laboratorium dan implementasi praktis, ada kebutuhan mendesak untuk pengumpulan data biometrik multimodal realistik yang memungkinkan untuk menyimpulkan hasil yang valid dari kondisi eksperimen terkontrol hingga aplikasi akhir.

Untuk mengatasi kelemahan signifikan dari sistem biometrik unimodal, biometrik multimodal telah menarik minat banyak peneliti[6]. Nageshkumar dkk.[7] memperkenalkan teknik otentikasi untuk sistem biometrik multimodal menggunakan 2 sifat yaitu wajah dan telapak tangan. Sistem tersebut telah dirancang untuk aplikasi di mana data pelatihan berisi wajah dan telapak tangan. Menggabungkan karakteristik wajah dan sidik jari dapat meningkatkan ketangguhan untuk mengotentikasi individu. Keputusan akhir dibuat dengan fusi pada tingkat skor pencocokan di mana vektor fitur dibangun secara independen untuk ukuran kueri dan kemudian dibandingkan dengan templat pendaftaran, yang disimpan selama persiapan basis data. Sistem biometrik multimodal ini dikembangkan melalui perpaduan pengenalan wajah dan sidik jari.

Berkaitan dengan Implementasiya menerapkan arsitektur jaringan saraf baru untuk klasifikasi multimodal biometrik yaitu dengan adanya penerapan jaringan *hierarki attention* yang mana pada jaringan ini dirancang untuk menangkap dua wawasan dasar dalam struktur dalam hal ini berkaitan database multimodal biometrik. Dalam kontribusi ini penulis mendapat gambaran Basis Data Multimodal BiosecurID yang diperoleh dalam proyek BiosecurID didukung oleh MEC Spanyol, dan dilakukan oleh konsorsium 6 Universitas Spanyol, Universidad Autonoma de Madrid (UAM), Universidad Politecnica de Madrid (UPM), Universidad Politecnica de Catalunya (UPC, Kampus Terrasa dan Kampus Mataro), Universidad de Zaragoza (UniZar), Universidad de Valladolid (UVA), dan Universidad del Pais Vasco (UPV)[5]. Tujuan utama dari Implementasi ini adalah akuisisi database multimodal yang realistik, yang secara statistik mewakili pengguna potensial aplikasi biometrik masa depan. Pada implementasi ini terdapat 4 Fitur yaitu gambar wajah, suara (Audio), tanda tangan, dan telapak tangan.

Berdasarkan permasalahan tersebut, maka penulis membuat implementasi jaringan *hierarki attention* untuk klasifikasi basis data multimodal biometrik suatu sistem untuk klasifikasi basis data multimodal biometrik dengan jaringan *hierarki attention* sehingga dapat menanggapi permasalahan yang telah ada sebelumnya dan menghasilkan performa atau akurasi yang tinggi.

II. METODE

Penulisan dimulai dengan studi literatur yang relevan dan kemudian dilanjutkan dengan pengumpulan data yang akan digunakan untuk membuat model berupa dataset multimodal. Selain itu, Pemrosesan data multimodal berdasarkan jenis datanya yaitu berupa Arsitektur *feature extraction*. Setelah itu, Arsitektur *Hierarki*, dan kemudian Evaluasi Matrik.

A. Dataset

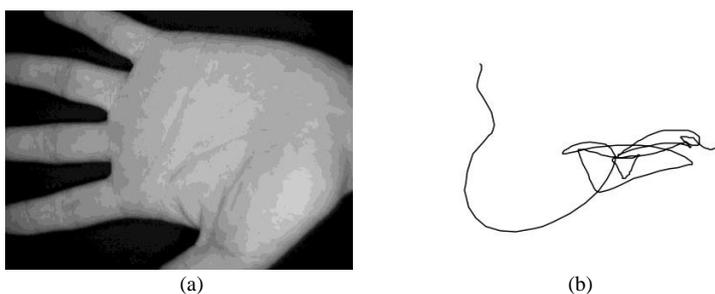
Dataset multimodal merupakan suatu kumpulan dataset yang terdiri dari informasi biometrik seseorang dengan jenis data seperti telapak tangan, wajah, audio, dan tanda tangan seseorang, namun dikarenakan adanya kesulitan dalam mengumpulkan dataset tersebut maka penulis berinisiatif membuat sebuah kombinasi dari beberapa dataset yang telah ada secara virtual sehingga dapat menjadi dataset prototipe untuk diterapkan pada arsitektur yang kami rancang. Berikut penulis paparkan sumber dataset yang digunakan dalam membuat dataset virtual.

Tabel 1. Dataset Multimodal yang digunakan

No.	Jenis Data	Sumber	Input Dimensi
1.	Wajah	Casia Face Database V.5	(1550, 160, 160, 1)
2.	Telapak Tangan	Casia Palmprint Database V.5	(1550, 32, 32, 1)
3.	Tanda Tangan	ATVS Database	(1550, 224, 224, 3)
4.	Audio	VoxCeleb1	(1550, 545, 13)

Dapat dilihat pada tabel 1 bahwa terdapat jenis data, sumber dan Input dimensi yang penulis gunakan, adapun Jenis Data terdiri dari Wajah, Telapak Tangan, Tanda Tangan, dan Audio. Jenis Data Wajah bersumber dari Casia Face Database V.5, input dimensi yang digunakan yaitu (1550, 160, 160, 1), jenis data telapak tangan bersumber dari Casia Palmprint Database V.5, Input Dimensi yang digunakan yaitu (1550, 32, 32, 1), jenis data Tanda tangan bersumber dari ATVS Database, input dimensi yang digunakan yaitu (1550, 224, 224, 3), dan jenis data Audio bersumber dari VoxCeleb1, dan input dimensi yang digunakan yaitu (1550, 545, 13).

Pada data jenis wajah dan telapak tangan merupakan sebuah data dengan tipe berupa gambar yang keduanya memiliki ukuran dimensi 640 x 480 x 3, kemudian pada data tanda tangan data original memiliki format *txt* dengan 5 kolom di dalamnya yakni koordinat X, koordinat Y, timestamp, kondisi pena, dan tekanan, berdasarkan 5 kolom tersebut penulis ubah ke dalam bentuk visualisasi gambar dengan dimensi 224 x 224 x 3, dan terakhir pada dataset audio yang memiliki tipe data jenis signal MFCC penulis menggunakan dimensi original nya yakni 545 x 13. Adapun sebagian dari beberapa dataset yang penulis gunakan sebagai berikut:



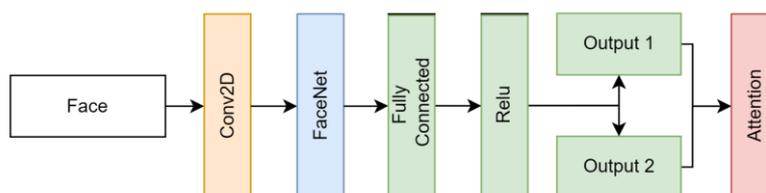
Gambar 1. Dataset Multimodal (a) Telapak tangan, (b) Tanda Tangan

Dapat dilihat pada gambar 1 yaitu (a) dan (b) merupakan dataset yang sebagian penulis gunakan dalam penelitian ini pada (a) adalah dataset telapak tangan dan pada (b) dataset tanda tangan.

B. Arsitektur *feature extraction*

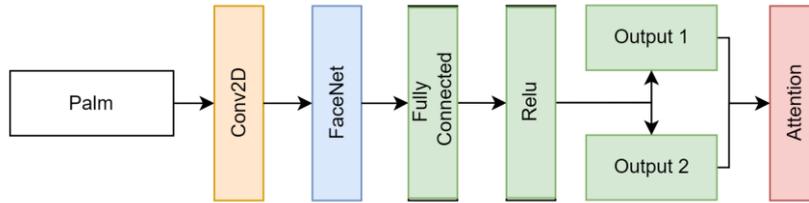
Berikut penulis paparkan penggunaan arsitektur untuk melakukan pemrosesan data multimodal berdasarkan jenis-jenis dataset nya masing-masing.

1) Wajah



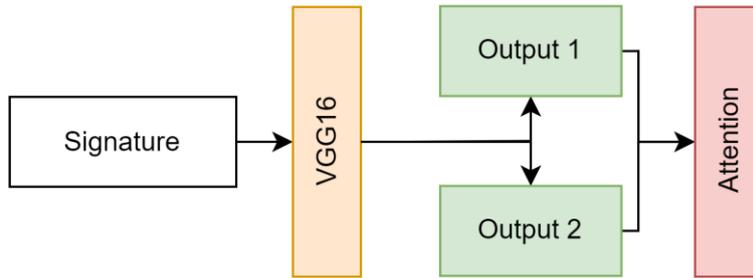
Gambar 2. Arsitektur *feature extraction* Wajah

Dapat dilihat pada gambar 2
 2) Telapak Tangan



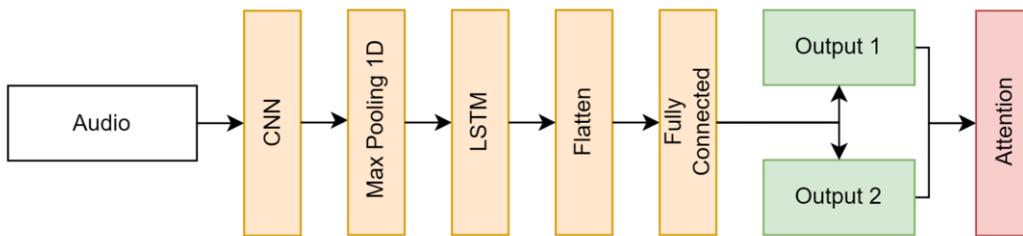
Gambar 3. Arsitektur *feature extraction* Telapak Tangan

Dapat dilihat pada gambar 3
 3) Tanda Tangan



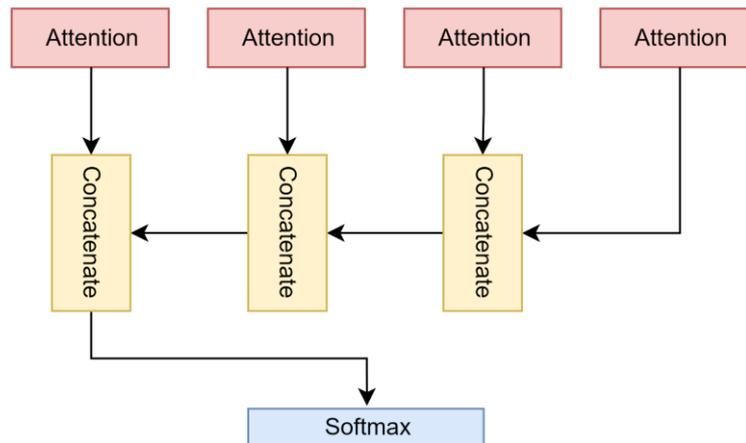
Gambar 4. Arsitektur *feature extraction* Tanda Tangan

Dapat dilihat pada gambar 4
 4) Audio



Gambar 5. Arsitektur *feature extraction* Audio

Dapat dilihat pada gambar 5
 C. Arsitektur Hierarki



Gambar 6. Arsitektur Hierarki yang digunakan

Dapat dilihat pada gambar 6

Tabel 2. Parameter dan Value yang digunakan

No.	Parameter	Value
1	Optimizer	Adam, learning_rate = 0.001

2	Loss	Categorical Cross Entropy
3	Epoch	50, stop at 25.
4	Total Parameter	52,097,586
5	Callbacks	Custom Callbacks(acc = 0.99)

Dilihat dari tabel 2 bahwa terdapat parameter dan value yang digunakan pada penelitian ini, Arsitektur Hierarki ini terdiri dari 4 *Attention*, 3 *Concatenate*, dan 1 *Softmax*, dengan parameter dan *Value* yang digunakan *Optimizer* dengan *value Adam*, kemudian untuk loss menggunakan *Categorical Cross Entropy*, dengan *training data* yaitu 50 Epoch berhenti di 25, selanjutnya total parameter 50,097,586 dan terakhir untuk *Callbacks* dilakukan *custom* dengan *acc =0,99* atau 99 %.

D. Evaluasi Matrik

Evaluasi metrik perlu dilakukan untuk menemukan kombinasi model yang terbaik. Evaluasi metrik ini dilakukan dengan mempertimbangkan beberapa hal yaitu hasil akurasi dan *confusion matrik*. *Confusion matrix* adalah alat evaluasi visual yang digunakan umumnya digunakan dalam *machine learning* maupun *deep learning*. Kolom *confusion matrix* mewakili hasil kelas prediksi dan baris mewakili hasil kelas sebenarnya sehingga dapat menghitung semua kemungkinan kasus masalah klasifikasi. Dalam *confusion matrix* terdapat beberapa jenis yaitu *presisi*, *recall*, dan *f1-score*. Akurasi adalah rasio prediksi benar terhadap keseluruhan data. Presisi adalah perbandingan nilai prediksi benar positif dibandingkan dengan total hasil dengan prediksi positif. Recall adalah perbandingan nilai prediksi benar positif dengan seluruh data yang benar positif. *F1-Score* merupakan perbandingan rata-rata nilai presisi dan recall yang dibebankan. Akurasi dan *confusion matrix* dapat dibentuk seperti pada Persamaan (1), (2), (3), dan (4).

$$\text{Akurasi} = \frac{TP + TN}{\text{Total}} \quad (1)$$

$$\text{Presisi} = \frac{TP}{FP + TP} \quad (2)$$

$$\text{Recall} = \frac{TP}{FN + TP} \quad (3)$$

$$F1 - \text{Score} * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

III. HASIL DAN PEMBAHASAN

A. Kegunaan tiap-tiap layer *feature extraction*

Pada *feature extraction* yang penulis gunakan disini menggunakan FaceNet, Mobilenet, dan CNN + LSTM. Sesuai dengan kegunaan masing-masing. Pada FaceNet merupakan sistem pengenalan wajah yang dikembangkan oleh peneliti Google. FaceNet mengekstrak fitur wajah menjadi vektor menggunakan arsitektur *deep Convolutional Neural Network* (deep CNN). Vektor nilai atau vector embedding yang dihasilkan dapat memetakan kemiripan wajah yang memiliki kedekatan posisi pada *embedding space*. Model Deep CNN yang digunakan pada FaceNet bisaberupa ZF-Net atau *Inception*[8]. Keuntungan dari model ini adalah hanya sedikit pemrosesan pada citra yang dapat digunakan sebagai input. Pada saat yang sama, akurasi model sangat tinggi dalam kumpulan data. FaceNet dapat digunakan secara luas dalam pengenalan wajah di terminal seluler.

Jaringan FaceNet terdiri dari lapisan input batch dan jaringan konvolusi yang dalam, dan kemudian normalisasi L2, yang mengarah ke penyematan wajah, dan akhirnya menghitung kehilangan triplet untuk membuat jarak antara objek yang sama[9]. Sekecil mungkin, jarak antara objek yang berbeda adalah sebesar mungkin. Itu menggunakan jaringan saraf konvolusi yang dalam untuk mempelajari metode penyisipan Euclidean dari setiap gambar, dan melatih jaringan sehingga jarak kuadrat L2 dalam ruang penyematan secara langsung sesuai dengan kemiripan wajah. FaceNet langsung menggunakan fungsi Loss dari LMNN (*Maximum Boundary Nearest Neighbor Classification*) berbasis Triplet untuk melatih jaringan saraf, dan keluaran jaringan adalah ruang vektor 128-dimensi. Triplet yang dipilih berisi dua gambar mini wajah yang cocok dan gambar mini wajah yang tidak cocok. Target fungsi Loss membedakan kelas positif dan negatif dengan batas jarak. Sehingga dalam hal ini untuk ekstraksi fitur pada wajah dapat dengan mudah pengaplikasiannya[9].

MobileNet merupakan salah satu arsitektur Convolutional Neural Network (CNN) yang dapat digunakan untuk mengatasi kebutuhan akan computing data dalam jumlah yang besar[10]. Perbedaan mendasar antara arsitektur MobileNet dan arsitektur CNN pada umumnya adalah penggunaan lapisan atau layer konvolusi dengan ketebalan filter yang sesuai dengan ketebalan dari input image[11]. MobileNet membagi konvolusi menjadi *depthwise convolution* dan *pointwise convolution*. MobileNet adalah salah satu yang paling canggih arsitektur yang secara signifikan dapat mengurangi ukuran model dan hanya mempengaruhi akurasi minima[11]. Dasar dari Arsitektur MobileNet adalah dua jenis: lapisan konvolusi yang berbeda dari standar proses konvolusi. Lapisan-lapisan ini dapat dipisahkan secara mendalam konvolusi dan lapisan konvolusi *pointwise*. Itu lapisan *depthwise* menggunakan filter tunggal untuk setiap saluran input, berbeda dari konvolusi standar yang menggunakan input dengan dimensi saluran untuk satu filter. Untuk lapisan titik, kemudian lakukan operasi perkalian filter 1x1 untuk digabungkan hasil peta fitur dari lapisan kedalaman[12].

CNN merupakan jenis *deep learning* yang memanfaatkan convolutional layer sebagai penyusun neural network yang dibangun. Banyak penelitian yang membuktikan bahwa CNN lebih baik daripada metode lain dalam melakukan pengolahan teks. Salah satunya dalam arsitektur VGG 16 yang terdiri dari 16 layer yang membuat performa model menjadi lebih baik karena semakin dalam layer. Telah dinyatakan bahwa convolutional layer pada CNN dapat mengurangi beban komputasi dari *neural network*. Hal ini terjadi karena *convolutional layer* pada dasarnya adalah sebuah *sparse matrix* yang dimensinya lebih kecil dari dimensi data yang diolah[13].

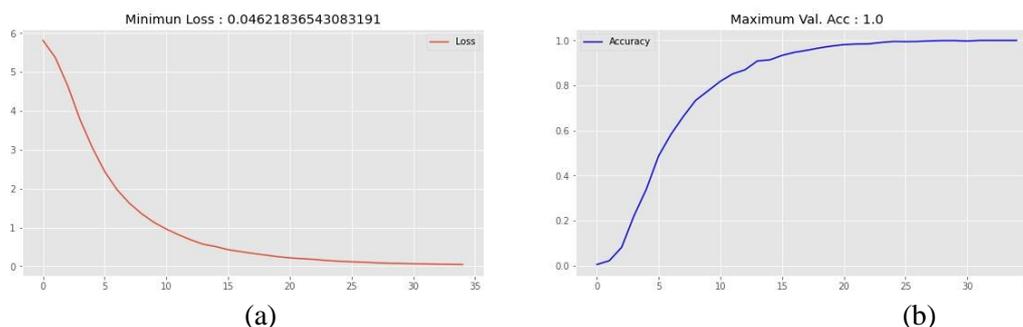
LSTM (*Long Short-Term Memory*) merupakan pengembangan dari metode RNN melalui penambahan sel (*cell*) LSTM di dalam arsitektur RNN. LSTM telah sukses menyelesaikan berbagai permasalahan, seperti *handwriting recognition*, *speech recognition*, *hand writing generation*, dan *image captioning*. LSTM memungkinkan arsitektur machine learning menyimpan bobot (*weight*) dari suatu perhitungan lebih lama dari RNN. Hal ini disebabkan LSTM memiliki sel-sel LSTM, yaitu sebuah node yang memiliki *self-recurrent*[13]. Hal ini menyebabkan LSTM dapat bekerja lebih baik daripada RNN pada data dengan sekuens yang lebih panjang. Pada LSTM, proses *recurrent* atau perulangan dilakukan pada level node dan layer. Sel-sel LSTM dikendalikan oleh input gate untuk mengingat atau melupakan informasi yang dimiliki berdasarkan keluaran dari LSTM. Pada masing-masing *Feature Extraction* yang digunakan mempunyai keunggulan masing-masing dan penting dalam memilihnya sesuai apa yang dibuat.

Pada masing-masing *Feature Extraction* yang digunakan memiliki keunggulan masing-masing untuk menghasilkan performa model yang baik dalam hal ini sesuai dengan jenis dataset biometrik multimodal yang digunakan yaitu telapak tangan, wajah, audio, dan tanda tangan seseorang memudahkan dalam proses yang akan dilakukan dengan sesuai dengan kasus yang dilakukan sehingga perlunya *Feature Extraction* dari *Facenet*, *MobileNet*, dan *CNN + LSTM* berbeda sesuai dataset yang ada dengan memperhatikan cara kerjanya.

B. Kegunaan Hierarki Attention

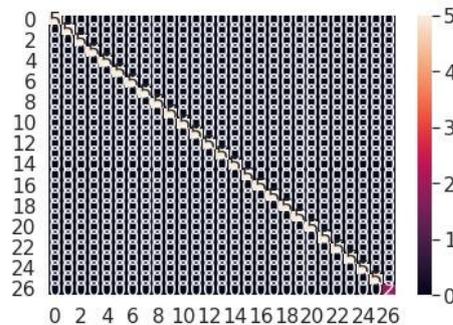
Penulis merumuskan jaringan arsitektur *Hierarki Attention* berdasarkan kombinasi dataset virtual dari *Casia Face Database V.5*, *Casia Palmprint Database V.5*, *ATVS Database* dengan tujuan untuk mengimprovisasi teknik identifikasi pengguna pada suatu sistem berdasarkan kombinasi data biometric sehingga celah keamanan sistem pengenalan lebih sulit untuk diretas, berdasarkan hasil dari pelatihan jaringan arsitektur *Hierarki Attention* yang menggabungkan arsitektur Ekstraksi Fitur seperti *Pretrained Model MobileNet*, *FaceNet Embedding*, dan *CNN-LSTM* menghasilkan feature vector yang unik pada setiap label sehingga arsitektur *Hierarki* dapat melakukan kombinasi *feature vector* secara bertingkat dengan dilengkapi arsitektur *Attention* sehingga *feature vector* hasil kombinasi dapat di transformasi menjadi lebih efisien dan mengurangi nilai *feature vector* yang menurun akibat lapisan *Fully Connected Layer*. Akurasi

Hasil akurasi yang didapatkan dalam penelitian ini bahwa menghasilkan hasil akurasi penulis ampirkan sebagai berikut:



Gambar 7. Hasil Akurasi (a) *Validation Loss*, (b) *Validation Accuracy*

Dapat dilihat pada gambar 7 terdapat dua buah grafik garis yang menunjukkan *loss* (grafik sebelah kiri) atau (a) dan *accuracy* (grafik sebelah kanan) atau (b) dari model yang telah didapatkan. Terlihat pada grafik sebelah kiri *validation loss*-nya didapatkan minimum *loss* 0,046 dan grafiknya naik. Pada grafik sebelah kanan yang menunjukkan *accuracy*, *validation accuracy* dimulai dari sangat kecil lalu perlahan lahan berhasil mendekati *accuracy* dan puncak terbaiknya menghasilkan *Maximum Validation Accuracy* sebesar 1.0 atau 100%.



Gambar 8. Hasil Confusion Matrik

Dapat dilihat pada gambar 8 bahwa pada hasil confusion matrik yang didapatkan dalam penelitian ini dari hasil training didapatkan masing-masing kelas mulai dari Akurasi, Presisi, *Recall*, *F1-Score* sebesar 1.0 atau 100% dapat dilihat dalam Gambar 8 bahwa dalam gradasi warna yang terdapat dari hasil confusion matrik dengan hasil yang didapatkan baik, karena dalam skala dari 0 sampai 5 atau dari skala kecil sampai yang tertinggi, didapatkan di skala 5 yang berarti model baik dan performa model yang dihasilkan baik.

IV. KESIMPULAN

Berdasarkan hasil penelitian dapat diambil kesimpulan bahwa dalam implementasi jaringan *hierarki attention* untuk klasifikasi basis data multimodal biometrik dalam implementasinya pada feature extraction pada setiap kelas Biometrik Multimodal dari dataset dengan jenis data telapak tangan, wajah, audio, dan tanda tangan seseorang berbeda-beda sesuai dengan keperluan yang dibutuhkan sehingga menghasilkan model yang baik sekaligus menerapkan jaringan *hierarki attention* dengan kombinasi dengan struktur hierarki menghasilkan akurasi sebesar 1.0 atau 100% sehingga dikatakan model yang didapatkan atau performa yang didapatkan baik diterapkan.

DAFTAR PUSTAKA

- [1] P. A. Nugroho, I. Fenriana, R. Arijanto, and M. Kom, "Implementasi Deep Learning Menggunakan Convolutional Neural Network (Cnn) Pada Ekspresi Manusia," *Implementasi Deep Learn. Menggunakan Convolutional Neural Netw. (Cnn) Pada Ekspresi Mns.*, vol. 2, no. 1, 2020.
- [2] N. Alay and H. H. Al-Baity, "Deep learning approach for multimodal biometric recognition system based on fusion of iris, face, and finger vein traits," *Sensors (Switzerland)*, vol. 20, no. 19, pp. 1–17, 2020, doi: 10.3390/s2019523.
- [3] L. Chato and S. Latifi, "Application of Machine Learning to Biometric Systems- A Survey," *J. Phys. Conf. Ser.*, vol. 1098, no. 1, 2018, doi: 10.1088/1742-6596/1098/1/012017.
- [4] A. A. Andarinny, C. E. Widodo, and K. Adi, "Perancangan sistem identifikasi biometrik jari tangan menggunakan Laplacian of Gaussian dan ekstraksi kontur," *Youngster Phys. J.*, vol. 6, no. 4, pp. 304–314, 2017.
- [5] J. Fierrez *et al.*, "BiosecurID: A multimodal biometric database," *Pattern Anal. Appl.*, vol. 13, no. 2, pp. 235–246, 2010, doi: 10.1007/s10044-009-0151-4.
- [6] A. A. E. Wahyudi, I. M. O. Widyantara, and K. O. Saputra, "Verifikasi Telapak Tangan dengan Multimodal Biometrik Fermat's Spiral," *J. RESTI (Rekayasa Sist. dan Teknol. Informatika)*, vol. 3, no. 3, pp. 408–413, 2019, doi: 10.29207/resti.v3i3.1173.
- [7] A. A. Ross, K. Nandakumar, and A. K. Jain, *Handbook of Multiometrics*, Internatio. Springer AS, 2006.
- [8] Miftakhurrokhmat, R. A. Rajagede, and R. Rahmadi, "Presensi Kelas Berbasis Pola Wajah, Senyum dan Wi-Fi Terdekat dengan Deep Learning," *J. RESTI (Rekayasa Sist. dan Teknol. Informatika)*, vol. 5, no. 1, pp. 31–38, 2021, doi: 10.29207/resti.v5i1.2575.
- [9] X. Xu, M. Du, H. Guo, J. Chang, and X. Zhao, "Lightweight FaceNet Based on MobileNet," *Int. J. Intell. Sci.*, vol. 11, no. 01, pp. 1–16, 2021, doi: 10.4236/ijis.2021.111001.
- [10] I. A. Dewi, "Deteksi Manusia menggunakan Pre-Trained MobileNet untuk Segmentasi Citra Menentukan Bentuk Tubuh," *MIND J.*, vol. 1, no. 2, pp. 65–79, 2019, doi: 10.26760/mindjournal.v4i1.65-79.
- [11] Hendriyana and Y. H. Maulana, "Identifikasi Jenis Kayu Menggunakan Convolutional Neural Network Dengan Arsitektur Mobilenet," *Ristek*, vol. 4, no. 1, pp. 70–76, 2020, [Online]. Available: <http://jurnal.iaii.or.id/index.php/RESTI/article/view/1445/203>.
- [12] Y. Harjoseputro, I. P. Yuda, and K. P. Danukusumo, "MobileNets: Efficient Convolutional Neural Network for Identification of Protected Birds," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 10, no. 6, pp. 2290–2296, 2020, doi: 10.18517/ijaseit.10.6.10948.
- [13] Yudi Widhiyana, Transmissia Semiawan, Ilham Gibran Achmad Mudzakir, and Muhammad Randi Noor, "Penerapan Convolutional Long Short-Term Memory untuk Klasifikasi Teks Berita Bahasa Indonesia," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 10, no. 4, pp. 354–361, 2021, doi: 10.22146/jnteti.v10i4.2438.