



Prediksi Kepribadian OCEAN dari Video Wajah dengan ArcFace dan LSTM

OCEAN Personality Prediction from Facial Videos Using ArcFace and LSTM

Talhah Alfian Padma, Salamun Rohman Nudin*

Fakultas Vokasi, Program Studi D4 Manajemen Informatika, Universitas Negeri Surabaya, Surabaya, Indonesia

Email: ¹talhahalfian.21054@mhs.unesa.ac.id, ²*salamunrohman@unesa.ac.id |

Email penulis korespondensi: salamunrohman@unesa.ac.id

Received: 27/09/2026, Revised: 10/10/2026, Accepted: 14/10/2026, Published: 23/10/2026

Abstrak—Penilaian kepribadian berbasis kuesioner dapat dipengaruhi bias laporan diri, sedangkan satu citra wajah tidak merekam perubahan perilaku selama interaksi. Karena itu, pemodelan video wajah menjadi pendekatan komplementer untuk mengukur kesan kepribadian yang tampak, bukan pengganti asesmen psikologis. Penelitian ini mengevaluasi prediksi kontinu lima dimensi Big Five (Openness, Conscientiousness, Extraversion, Agreeableness, dan Neuroticism) dengan transfer learning ArcFace dan Long Short-Term Memory (LSTM). Dataset yang digunakan adalah ChaLearn Looking at People First Impressions V2 yang terdiri atas 10.000 klip; sepuluh frame dipilih merata dari tiap video, wajah dideteksi menggunakan MTCNN, lalu dipotong dan diubah ke ukuran 112×112 piksel. Backbone ArcFace berbasis ResNet34 menghasilkan embedding 512 dimensi untuk setiap frame, sedangkan dua lapisan LSTM merangkum urutan temporal sebelum lapisan regresi multi-output memprediksi lima skor dalam rentang 0–1. Evaluasi menggunakan Mean Absolute Error (MAE); persentase yang dilaporkan dihitung sebagai $(1 - \text{MAE}) \times 100$ dan disebut akurasi regresi, sehingga tidak disamakan dengan akurasi klasifikasi. Pada data validasi, RMSprop memperoleh 90,52%; pada data uji, konfigurasi yang diberi label baseline memperoleh 90,46% dan RMSprop 90,42%. Pada RMSprop, Agreeableness mencapai 90,85%, sedangkan Neuroticism 89,84%. Hasil ini menunjukkan bahwa representasi wajah berurutan dapat mendukung estimasi kesan kepribadian dalam benchmark tersebut, dengan selisih antarkonfigurasi yang kecil. Secara praktis, sistem dapat menjadi alat bantu analisis perilaku awal dalam konteks terkontrol, tetapi skor wajah saja tidak memadai untuk keputusan individual berisiko tinggi. Replikasi dan validasi lintas populasi perlu dilakukan sebelum penggunaan di luar benchmark.

Kata Kunci: Big Five Personality; Transfer Learning; ArcFace; Prediksi Kepribadian; RMSprop

Abstract—Questionnaire-based personality assessment can be affected by self-report bias, while a single facial image cannot represent behavioral change during an interaction. Video-based facial modeling is therefore a complementary way to estimate apparent personality, rather than a substitute for psychological assessment. This study evaluates continuous prediction of the Big Five dimensions (Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism) using ArcFace transfer learning and Long Short-Term Memory (LSTM). The ChaLearn Looking at People First Impressions V2 dataset contains 10,000 clips. Ten frames were sampled uniformly from each video, faces were detected with MTCNN, and crops were resized to 112×112 pixels. A ResNet34-based ArcFace backbone produced a 512-dimensional embedding for each frame. Two LSTM layers summarized the temporal sequence, and a multi-output regression head predicted five scores in the range 0–1. Mean Absolute Error (MAE) was used for evaluation; the reported percentage was computed as $(1 - \text{MAE}) \times 100$ and is termed regression accuracy, not classification accuracy. RMSprop achieved 90.52% on validation, while the configuration labeled baseline achieved 90.46% and RMSprop achieved 90.42% on the test split. With RMSprop, Agreeableness reached 90.85%, whereas Neuroticism reached 89.84%. These results indicate that sequential facial representations can support apparent personality estimation on this benchmark, although the differences between configurations are small. In practice, the system may support preliminary behavioral analysis in controlled settings, but facial scores alone are insufficient for high-stakes individual decisions. Replication and cross-population validation are needed before use beyond the benchmark.

Keywords: Big Five Personality; Transfer Learning; ArcFace; Personality Prediction; RMSprop

1. PENDAHULUAN

Kepribadian merangkum kecenderungan perilaku, emosi, dan cara individu berinteraksi; model lima faktor OCEAN menyediakan lima dimensi kontinu untuk mendeskripsikan perbedaan tersebut [1], [2]. Dalam penelitian komputasional, sasaran yang diprediksi dari video wawancara sering kali merupakan apparent personality, yaitu kesan yang dibentuk pengamat dari perilaku yang tampak, bukan keseluruhan sifat internal seseorang. Perbedaan ini penting label crowdsourced pada ChaLearn First Impressions V2 merekam penilaian pengamat terhadap klip, sehingga skor model harus ditafsirkan sebagai estimasi kesan berbasis benchmark, bukan diagnosis atau pengukuran psikometrik langsung. Kuesioner tetap memberi akses pada laporan diri yang tidak terlihat dari wajah, tetapi jawabannya rentan terhadap bias ingatan, pemahaman item, dan keinginan memberi kesan sosial yang baik. Sementara itu, penilaian visual dapat diskalakan untuk menganalisis banyak klip, namun juga membawa bias pengamat dan konteks rekaman. Dengan demikian, motivasi penelitian ini adalah menguji representasi visual temporal sebagai sumber bukti komplementer dan terukur, bukan menyatakan bahwa ekspresi wajah menentukan kepribadian.

Penelitian terdahulu menunjukkan kemajuan tetapi juga batas yang berbeda dengan menggabungkan fitur visual wajah dan suasana, audio, serta teks dengan CNN dan beberapa strategi fusi kekuatan pendekatan tersebut adalah pemanfaatan sinyal lintas modalitas, sedangkan ketergantungannya pada audio dan transkrip membuat perbandingan tidak setara ketika penelitian hanya memakai wajah. Artikel itu juga menunjukkan bahwa sejumlah kecil frame dapat memberi hasil yang sebanding, sehingga pemilihan frame menjadi faktor penting dalam biaya komputasi [3]. Penelitian terdahulu





menggabungkan fitur audio dan visual dengan arsitektur hibrida pada First Impressions V2 integrasi berpotensi menambah konteks prosodi, tetapi menambah tahapan ekstraksi dan kebutuhan sinkronisasi yang tidak tersedia dalam pendekatan visual tunggal [4]. Terdapat penelitian terdahulu juga membandingkan fitur modalitas individual dan fusi, melaporkan bahwa fitur adegan/visual dan fusi awal dapat menghasilkan rerata kuat pada split ChaLearn; hasil tersebut menegaskan bahwa jenis fitur, rancangan fusi, dan protokol split memengaruhi angka, sehingga satu persentase tidak dapat dibandingkan tanpa menyamakan setup [5]. memakai klip untuk meningkatkan efektivitas dan penjelasan apparent personality, namun representasi visi-bahasa bergantung pada pretraining skala besar dan rancangan prompt/fitur yang berlainan dari ekstraksi embedding wajah [6]. Adapula secara khusus menguji fitur emosi per frame dengan LSTM dan fitur wajah statis, akurasi yang dilaporkan 90,67% menunjukkan kontribusi potensial dinamika ekspresi tetapi sistem mereka juga memakai demografi dan definisi akurasi serta konfigurasi yang perlu diperiksa sebelum disejajarkan dengan studi ini [7]. Secara komparatif, metode multimodal dapat menangkap lebih banyak petunjuk namun lebih sulit direplikasi bila setiap modalitas tidak tersedia CNN statis efisien tetapi melewatkan urutan perubahan; model temporal berfokus emosi lebih relevan bagi dinamika, tetapi bergantung pada mutu pengenalan emosi dan dapat mereduksi ekspresi menjadi kategori-kategori terbatas. Berfokus pada mengulas ekstraksi fitur wajah untuk pengenalan trait, sedangkan penelitian lain mengestimasi Big Five dari transkripsi video kedua studi memperlihatkan ruang lingkup visual dan tekstual yang berbeda dari pipeline wajah temporal dalam penelitian ini [2], [5]. frame dapat memberi hasil yang sebanding, sehingga pemilihan frame menjadi faktor penting dalam biaya komputasi [4]. Penelitian lain menggabungkan fitur audio dan visual dengan arsitektur hibrida pada First Impressions V2; integrasi berpotensi menambah konteks prosodi, tetapi menambah tahapan ekstraksi dan kebutuhan sinkronisasi yang tidak tersedia dalam pendekatan visual tunggal [4]. Penelitian lainnya membandingkan fitur modalitas individual dan fusi, melaporkan bahwa fitur adegan/visual dan fusi awal dapat menghasilkan rerata kuat pada split ChaLearn; hasil tersebut menegaskan bahwa jenis fitur, rancangan fusi, dan protokol split memengaruhi angka, sehingga satu persentase tidak dapat dibandingkan tanpa menyamakan setup [3]. Penelitian lain memakai klip untuk meningkatkan efektivitas dan penjelasan apparent personality, namun representasi visi-bahasa bergantung pada pretraining skala besar dan rancangan prompt/fitur yang berlainan dari ekstraksi embedding wajah [6]. Terdapat penelitian terdahulu juga secara khusus menguji fitur emosi per frame dengan LSTM dan fitur wajah statis; akurasi yang dilaporkan 90,67% menunjukkan kontribusi potensial dinamika ekspresi, tetapi sistem mereka juga memakai demografi dan definisi akurasi serta konfigurasi yang perlu diperiksa sebelum disejajarkan dengan studi ini [7]. Secara komparatif, metode multimodal dapat menangkap lebih banyak petunjuk namun lebih sulit direplikasi bila setiap modalitas tidak tersedia; CNN statis efisien tetapi melewatkan urutan perubahan; model temporal berfokus emosi lebih relevan bagi dinamika, tetapi bergantung pada mutu pengenalan emosi dan dapat mereduksi ekspresi menjadi kategori-kategori terbatas. Terdapat penelitian terdahulu dengan pendekatan klasifikasi multi-task untuk prediksi Big Five [8].

Pemilihan ArcFace perlu dijelaskan dalam batas fungsi aslinya. ArcFace dirancang untuk pengenalan identitas wajah dan menerapkan additive angular margin pada ruang embedding ternormalisasi agar kelas identitas lebih terpisah secara sudut [9], [10], [11]. CNN konvensional seperti VGG-Face atau ResNet yang dilatih dengan objective klasifikasi umum menghasilkan fitur berguna, namun objective tersebut tidak secara eksplisit mengatur margin sudut identitas. Dalam studi ini, ArcFace bukan dipakai sebagai prediktor kepribadian dan tidak mengubah loss asalnya menjadi loss kepribadian; backbone beku berfungsi sebagai ekstraktor fitur frame, kemudian LSTM dan kepala regresi dipelajari terhadap label OCEAN. Keuntungannya adalah embedding wajah yang telah dipelajari dari skala data identitas besar dapat menjadi representasi awal yang relatif stabil, sedangkan LSTM mengintegrasikan perubahan antarframe. Kelemahannya adalah objective identitas dapat menekan variasi ekspresi yang tidak relevan bagi identitas, sehingga belum terbukti bahwa ArcFace unggul atas VGG-Face atau ResNet murni untuk regresi kepribadian. Klaim keunggulan komparatif memerlukan ablation study yang melatih backbone alternatif dengan split, preprocessing, seed, dan pencarian hiperparameter yang sama. Penelitian ini karenanya menguji kelayakan kombinasi ArcFace–LSTM dan membandingkan konfigurasi optimasi pada arsitektur yang sama, bukan mengklaim bahwa ArcFace secara universal lebih baik daripada CNN lain. ArcFace perlu dijelaskan dalam batas fungsi aslinya. ArcFace dirancang untuk pengenalan identitas wajah dan menerapkan additive angular margin pada ruang embedding ternormalisasi agar kelas identitas lebih terpisah secara sudut [10], [11]. CNN konvensional seperti VGG-Face atau ResNet yang dilatih dengan objective klasifikasi umum menghasilkan fitur berguna, namun objective tersebut tidak secara eksplisit mengatur margin sudut identitas. Dalam studi ini, ArcFace bukan dipakai sebagai prediktor kepribadian dan tidak mengubah loss asalnya menjadi loss kepribadian; backbone beku berfungsi sebagai ekstraktor fitur frame, kemudian LSTM dan kepala regresi dipelajari terhadap label OCEAN. Keuntungannya adalah embedding wajah yang telah dipelajari dari skala data identitas besar dapat menjadi representasi awal yang relatif stabil, sedangkan LSTM mengintegrasikan perubahan antarframe. Kelemahannya adalah objective identitas dapat menekan variasi ekspresi yang tidak relevan bagi identitas, sehingga belum terbukti bahwa ArcFace unggul atas VGG-Face atau ResNet murni untuk regresi kepribadian. Klaim keunggulan komparatif memerlukan ablation study yang melatih backbone alternatif dengan split, preprocessing, seed, dan pencarian hiperparameter yang sama. Penelitian ini karenanya menguji kelayakan kombinasi ArcFace–LSTM dan membandingkan konfigurasi optimasi pada arsitektur yang sama, bukan mengklaim bahwa ArcFace secara universal lebih baik daripada CNN lain.

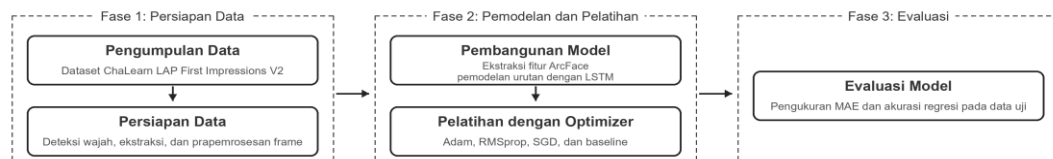
Secara teoritis, hubungan isyarat wajah dan OCEAN bersifat probabilistik dan kontekstual. Openness dapat berkaitan dengan variasi ekspresi dan keterlibatan yang ditampilkan saat membahas topik berbeda; Conscientiousness mungkin berkaitan dengan kestabilan perilaku ekspresif atau pengaturan diri, tetapi wajah saja tidak mengungkapkan kedisiplinan secara langsung. Extraversion sering dinilai melalui senyum, intensitas ekspresi, dan keterlibatan sosial yang tampak; Agreeableness dapat tercermin dalam isyarat afiliasi seperti senyum sosial dan ekspresi responsif; Neuroticism

dapat berkaitan dengan ketegangan atau perubahan afek negatif. Hubungan tersebut menggambarkan cue yang mungkin digunakan pengamat ketika membentuk kesan, bukan biomarker yang menetapkan sifat internal. Micro-expression secara definisi adalah gerak wajah sangat singkat dan beramplitudo rendah; sepuluh frame yang disampel merata dari klip tidak menjamin semua micro-expression tertangkap. Karena itu, naskah ini menggunakan istilah dinamika ekspresi temporal secara hati-hati dan tidak menyatakan bahwa model secara langsung mengidentifikasi micro-expression tanpa modul deteksi khusus. Kualitas label crowdsourced, pencahayaan, pose, budaya, usia, dan perilaku kamera juga dapat memengaruhi kaitan cue dengan skor. Kajian mutakhir tentang ekspresi wajah juga menekankan tantangan variasi data dan konteks pengamatan [12]. Kajian terbaru juga mengingatkan bahwa inferensi sifat batin dari wajah menimbulkan persoalan validitas dan dampak sosial; karena itu, keluaran model ini ditafsirkan sebagai kesan pada benchmark, bukan pengukuran karakter individual [13].

Berdasarkan celah tersebut, penelitian ini membangun pipeline ArcFace–LSTM pada ChaLearn First Impressions V2 untuk memprediksi lima label kontinu secara simultan. Wajah dipotong dari sepuluh frame yang disampel seragam, backbone ArcFace ResNet34 menghasilkan embedding 512 dimensi, dan dua LSTM memodelkan urutan sebelum dense layer memetakan representasi ke lima skor. Evaluasi memakai MAE serta transformasi 1–MAE, dengan hasil validasi dan pengujian dipisahkan. Kontribusi yang diuji adalah (1) integrasi ekstraksi embedding ArcFace beku dan pemodelan temporal LSTM pada label OCEAN, (2) perbandingan konfigurasi Adam, RMSprop, SGD, dan baseline sesuai implementasi, dan (3) analisis hasil tiap dimensi serta dinamika kurva. Interpretasi dibatasi pada benchmark dan setup yang tercatat; generalisasi ke populasi nyata maupun manfaat seleksi kerja tidak diasumsikan.

2. METODOLOGI PENELITIAN

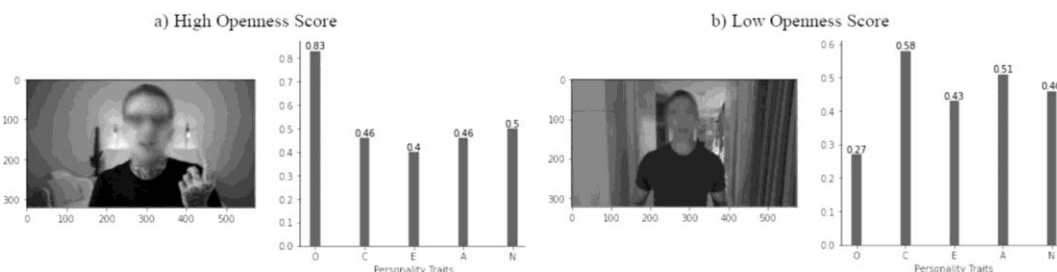
Gambar 1 merangkum alur penelitian dari pengumpulan ChaLearn First Impressions V2, pemilihan frame dan deteksi wajah, pembentukan embedding ArcFace, pemodelan temporal dengan LSTM, pelatihan regresi multi-output, hingga evaluasi pada split validasi dan pengujian. Alur tersebut menunjukkan letak penerapan solusi pada setiap tahap; rincian algoritma yang digunakan dijelaskan pada subbagian berikut [5], [14], [15].



Gambar 1. Tahap Penelitian

2.1. Pengumpulan Data

Data penelitian berupa dataset sekunder ChaLearn LAP First Impressions V2, yang merupakan pengembangan benchmark First Impressions [16], [17]. Dataset dibagi menjadi subset pelatihan, validasi, dan pengujian dengan rasio 3:1:1, masing-masing 6.000, 2.000, dan 2.000 video. Subset pelatihan digunakan untuk memperbarui parameter model, subset validasi untuk memantau kinerja selama pelatihan, dan subset pengujian untuk evaluasi akhir. Pemisahan ini menjaga agar evaluasi akhir dilakukan pada data yang tidak digunakan dalam pembaruan bobot. Contoh frame dan target skor yang digunakan sebagai masukan dan label diperlihatkan pada Gambar 2.



Gambar 2. Contoh Sampel Video Dataset ChaLearn LAP dengan Distribusi Skor Dimensi Kepribadian Big Five

Gambar 2 memperlihatkan contoh sekuens frame dari dataset ChaLearn LAP First Impressions V2 beserta skor ground truth untuk dimensi OCEAN. Contoh tersebut menunjukkan bahwa label kepribadian dinyatakan sebagai nilai kontinu pada rentang 0–1. Variasi visual antarsubjek menjadi masukan model, sedangkan skor anotasi menjadi target yang dipelajari.

2.2. Persiapan Data

Prapemrosesan mengubah video mentah menjadi sekuens wajah dengan dimensi yang konsisten untuk masukan model. Tahapan utamanya mencakup pemilihan frame, deteksi dan pemotongan wajah, pencocokan label, serta penyimpanan data dalam format yang mendukung pemuatan batch.

2.2.1 Ekstraksi Wajah dari Video

Video MP4 dibaca menggunakan OpenCV, kemudian 10 frame dipilih secara merata dari awal hingga akhir klip [18]. Pada setiap frame terpilih, MTCNN digunakan untuk mendeteksi wajah dan menentukan area pemotongan [14]. Wajah yang terdeteksi dipotong dan diubah ukurannya menjadi 112×112 piksel RGB, sesuai masukan backbone ArcFace. Jika deteksi berhasil pada kurang dari 10 frame, strategi temporal fallback yang digunakan dalam implementasi menduplikasi wajah valid terakhir agar bentuk sekuens tetap (10, 112, 112, 3). Tahap deteksi, pemotongan, dan penyesuaian ukuran wajah ditunjukkan pada Gambar 3.



Gambar 3. Contoh Hasil Prapemrosesan Wajah Menggunakan Algoritma MTCNN

Gambar 3 memperlihatkan hasil prapemrosesan wajah menggunakan MTCNN. Alur visualnya dimulai dari pemilihan frame, dilanjutkan dengan pelokalan wajah, lalu pemotongan dan penyesuaian ukuran citra sebelum ekstraksi fitur.

- Setiap klip dibaca menggunakan OpenCV dan diambil 10 frame secara merata berdasarkan posisi dalam durasi video. Pemilihan ini membentuk sekuens ringkas yang mewakili bagian-bagian video dari awal hingga akhir.
- MTCNN menjalankan deteksi wajah bertahap melalui P-Net, R-Net, dan O-Net untuk menghasilkan lokasi wajah dan landmark [13], [13]. Area wajah pada setiap frame terpilih kemudian digunakan sebagai masukan tahap berikutnya.
- Citra pada bounding box wajah dipotong dan diubah ukurannya menjadi 112×112 piksel RGB. Ukuran ini disesuaikan dengan masukan ArcFace. Landmark wajah membantu menjelaskan proses pelokalan pada ilustrasi; fitur yang digunakan model pada penelitian ini diekstraksi oleh backbone ArcFace.

2.2.2 Anotasi Data

Setiap video dipasangkan dengan lima label kontinu OCEAN yang disediakan dalam anotasi dataset. Identitas berkas digunakan untuk mencocokkan label dengan sekuens wajah hasil prapemrosesan. Pasangan input-target yang terbentuk berupa tensor wajah berukuran (10, 112, 112, 3) dan vektor label berdimensi lima dengan nilai dalam rentang 0–1.

2.2.3 Penyimpanan Dataset

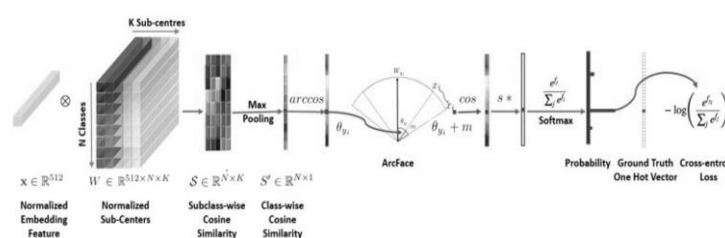
Sekuens wajah dan label yang telah dipasangkan disusun dalam pipeline tf.data.Dataset untuk pemuatan batch. Data terserialisasi disimpan agar dapat dibaca kembali selama pelatihan tanpa mengulang proses deteksi dan pemotongan wajah.

2.3. Pembangunan Model

Model terdiri atas ekstraktor fitur spasial ArcFace berbasis ResNet34 dan dua lapisan LSTM untuk pemodelan urutan temporal. Representasi keluaran kemudian diteruskan ke lapisan dense untuk menghasilkan lima skor OCEAN secara bersamaan.

2.3.1. Ekstraksi fitur Menggunakan Transfer Learning dengan ArcFace

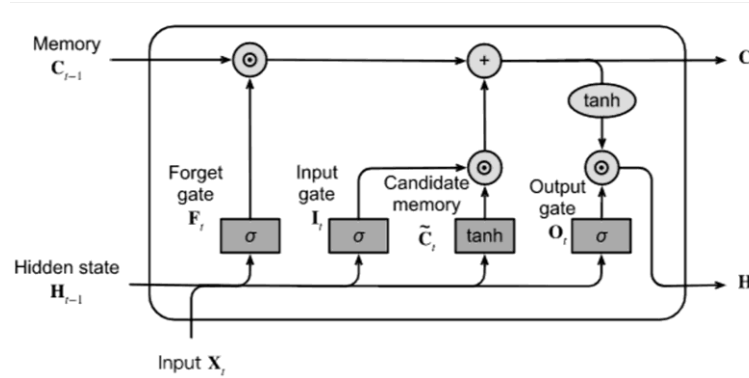
Backbone ArcFace dengan ResNet34 yang telah dilatih sebelumnya digunakan untuk menghasilkan embedding wajah berdimensi 512 pada setiap frame. Bobot backbone dibekukan selama pelatihan pada dataset target, sehingga pembaruan parameter berfokus pada bagian temporal dan lapisan prediksi. ArcFace dilatih menggunakan Additive Angular Margin Loss untuk membentuk representasi yang diskriminatif pada tugas pengenalan wajah [19]. Dalam penelitian ini, embedding tersebut dimanfaatkan sebagai fitur masukan; performa prediksi kepribadian tetap ditentukan melalui evaluasi pada dataset ChaLearn LAP. Arsitektur ekstraksi embedding ArcFace yang diterapkan dalam penelitian ditampilkan pada Gambar 4.



Gambar 4. Arsitektur ArcFace Model

2.3.2. LSTM untuk Analisis Temporal

Long Short-Term Memory (LSTM) adalah jenis Recurrent Neural Network yang menggunakan input gate, forget gate, dan output gate untuk mengatur informasi yang diteruskan sepanjang sekuens. Mekanisme ini membantu mempertahankan informasi relevan dari langkah sebelumnya ketika memproses data berurutan [20]. Pada model ini, LSTM memproses urutan embedding ArcFace dari 10 frame. Lapisan pertama mempertahankan keluaran sepanjang timestep, dan lapisan kedua merangkum urutan menjadi representasi temporal berdimensi 128 untuk diteruskan ke lapisan prediksi [15]. Struktur LSTM yang mengolah urutan embedding ditampilkan pada Gambar 5.



Gambar 5. Arsitektur Long Short-Term Memory

2.4. Optimizer dan Parameter Pelatihan

Eksperimen dijalankan pada komputer dengan prosesor Intel Core i9, RAM 32 GB, SSD 1 TB, dan GPU NVIDIA GeForce RTX 4080 dengan memori 16 GB. Pipeline pemuatan data menggunakan TensorFlow 2.13.0 melalui `tf.data.Dataset`[21]. Empat konfigurasi dibandingkan pada arsitektur ArcFace-LSTM dan pembagian data yang sama, dengan learning rate 0,001, batch size 8, dan maksimum 100 epoch. Early stopping terlihat pada kurva beberapa konfigurasi. Baseline didefinisikan sebagai konfigurasi yang menggunakan optimizer default framework tanpa hyperparameter tuning, dengan arsitektur ArcFace-LSTM tetap sama; baseline bukan model tanpa optimizer dan bukan model tanpa LSTM. Nama optimizer dan nilai default persisnya tidak tercatat, sehingga keduanya perlu dicantumkan dari kode eksperimen untuk replikasi penuh. Formulasi pembandingan SGD, RMSprop, dan Adam merujuk pada [22],[23].

2.4.1 Stochastic Gradient Descent (SGD)

SGD merupakan bentuk dasar dari algoritma pembaruan bobot, di mana parameter diperbarui secara linier berdasarkan turunan fungsi rugi (loss) terhadap parameter model dengan learning rate konstan [22].

Rumus pembaruan bobot SGD dirumuskan pada Persamaan (1):

$$W_{t+1} = W_t - \alpha \cdot \nabla Q_i(W_t) \quad W_{t+1} W_t \alpha \nabla Q_i(W_t) - i \quad (1)$$

Keterangan: adalah bobot baru, adalah bobot lama, adalah learning rate, dan) adalah gradien loss function terhadap bobot pada sampel ke.

2.4.2 Root Mean Square Propagation (RMSprop)

RMSprop adalah algoritma optimasi adaptif yang membagi gradien dengan akar rata-rata bergerak kuadrat gradien historis (root mean square of recent gradients), sehingga menstabilkan laju pembaruan bobot pada jaringan rekuren [17].

Rumus pembaruan bobot RMSprop dirumuskan pada Persamaan (2) dan (3):

$$v_{t+1} = \gamma \cdot v_t + (1 - \gamma) \cdot (\nabla f(W_t))^2 \quad (2)$$

$$W_{t+1} = W_t - [\alpha / (\sqrt{v_{t+1}} + \epsilon)] \cdot \nabla f(W_t) \quad (3)$$

Keterangan: adalah akumulasi rata-rata bergerak kuadrat gradien, adalah faktor peluruhan (decay factor), adalah global learning rate.

2.4.3 Adaptive Moment Estimation (Adam)

Adam menggabungkan konsep dari Momentum dan RMSprop dengan menyimpan rata-rata bergerak momen pertama (mean gradien) dan momen kedua (varians gradien) yang dikoreksi terhadap bias [18]. Rumus pembaruan bobot Adam dirumuskan pada Persamaan (4):

$$W_{t+1} = W_t - [\alpha / (\sqrt{\hat{v}_t} + \epsilon)] \cdot \hat{m}_t \quad (4)$$

Keterangan: $\hat{m}_t$ dan $\hat{v}_t$ adalah estimator bias-corrected untuk momen pertama dan momen kedua gradien.



2.5. Evaluasi Model

Evaluasi menggunakan Mean Absolute Error (MAE), yaitu rata-rata selisih absolut antara skor prediksi dan label aktual. MAE sesuai untuk mengukur besar kesalahan pada target kontinu dan dapat dihitung pada setiap dimensi maupun secara agregat. Untuk memudahkan pembacaan hasil, penelitian ini juga menyajikan transformasi MAE menjadi persentase akurasi regresi sesuai Persamaan (6); karena itu, nilai persentase tersebut dibaca bersama definisi transformasinya, bukan sebagai akurasi klasifikasi. Rumus perhitungan MAE dirumuskan pada Persamaan (5):

$$MAE = (1/n) \cdot \sum |Y_i - X_i| (i = 1s.d.n) \quad (5)$$

Keterangan: X_i adalah nilai ground truth aktual, Y_i adalah nilai prediksi model, dan n adalah jumlah sampel data. Nilai MAE yang lebih rendah menunjukkan performa prediksi yang lebih baik. Untuk menginterpretasikan MAE dalam bentuk persentase akurasi regresi, digunakan transformasi Persamaan (6):

$$Accuracy_{reg} = (1 - MAE) \times 100 \quad (6)$$

Prediksi dibandingkan dengan label ground truth pada subset pelatihan, validasi, dan pengujian untuk setiap konfigurasi optimizer. Analisis meliputi MAE dan persentase akurasi regresi secara keseluruhan, hasil per dimensi OCEAN, serta pola kurva pelatihan dan validasi.

3. HASIL DAN PEMBAHASAN

Bagian ini melaporkan hasil evaluasi model ArcFace-LSTM dalam memprediksi lima skor OCEAN. Kinerja dibandingkan untuk konfigurasi Adam, SGD, RMSprop, dan baseline pada subset pelatihan, validasi, dan pengujian. Interpretasi hasil mengacu pada metrik yang digunakan dan nilai yang tercantum pada tabel serta kurva.

3.1. Hasil Evaluasi Perbandingan Optimizer Pada Model

Tabel 1 merangkum persentase akurasi regresi pada subset pelatihan, validasi, dan pengujian untuk setiap konfigurasi optimizer.

Tabel 1. Evaluasi Perbandingan Optimizer Pada Model ArcFace-LSTM

Optimizer	Akurasi Training (%)	Akurasi Validation (%)	Akurasi Testing (%)
Adam	91,57%	90,19%	90,16%
Baseline	92,74%	90,42%	90,46%
RMSprop	93,68%	90,52%	90,42%
SGD	89,10%	89,28%	89,27%

Pada Tabel 1, RMSprop mencatat nilai tertinggi pada subset validasi sebesar 90,52%, disusul baseline 90,42%, Adam 90,19%, dan SGD 89,28%. Pada subset pengujian, baseline memperoleh 90,46%, RMSprop 90,42%, Adam 90,16%, dan SGD 89,27%. Dengan demikian, RMSprop unggul pada validasi, sedangkan baseline memiliki nilai pengujian tertinggi. Selisih yang kecil antara nilai validasi dan pengujian pada RMSprop menunjukkan hasil yang berdekatan pada dua subset tersebut, tetapi tidak dengan sendirinya membuktikan generalisasi untuk populasi atau dataset lain. Perbandingan ini menunjukkan bahwa urutan kinerja bergantung pada subset evaluasi. Nilai pelatihan RMSprop yang lebih tinggi daripada validasi dan pengujian juga mengingatkan bahwa hasil pelatihan saja tidak cukup untuk menilai kemampuan pada data baru. Pembahasan berikutnya karena itu meninjau kinerja per dimensi dan pola kurva, bukan hanya satu nilai agregat.

3.2. Hasil Akurasi Per Dimensi Big Five Personality (OCEAN)

Untuk melihat variasi kinerja antartrait, hasil juga dianalisis secara terpisah pada Openness, Conscientiousness, Extraversion, Agreeableness, dan Neuroticism untuk ketiga subset data.

Tabel 2. Hasil Akurasi Tiap Dimensi OCEAN Optimizer Adam

Optimizer Adam	Openness	Conscientiousness	Extraversion	Agreeableness	Neuroticism
Training	91,65%	91,44%	92,15%	91,84%	90,77%
Validation	90,31%	90,18%	90,47%	90,91%	89,08%
Testing	90,36%	90,15%	90,38%	90,58%	89,33%

Tabel 2 menyajikan hasil untuk konfigurasi Adam. Pada subset pengujian, nilai berada pada rentang 89,33%–90,58%. Agreeableness memperoleh nilai tertinggi sebesar 90,58%, sementara Neuroticism memperoleh nilai terendah sebesar 89,33%. Perbedaan ini menggambarkan variasi hasil pada data dan konfigurasi yang diuji, bukan tingkat kesulitan trait yang berlaku universal. Nilai Neuroticism pada pengujian lebih rendah daripada empat dimensi lainnya dalam konfigurasi Adam, tetapi selisihnya tetap terbatas. Data ini belum cukup untuk menyimpulkan bahwa penyebabnya adalah karakter ekspresi tertentu; kesimpulan tersebut memerlukan analisis fitur dan validasi tambahan.

**Tabel 3.** Hasil Akurasi Tiap Dimensi OCEAN Optimizer SGD

Optimizer SGD	Openness	Conscientiousness	Extraversion	Agreeableness	Neuroticism
Training	89,45%	88,82%	89,58%	89,65%	88,00%
Validation	89,52%	89,01%	89,66%	89,92%	88,29%
Testing	89,55%	88,98%	89,62%	89,88%	88,32%

Tabel 3 menyajikan hasil konfigurasi SGD. Akurasi pengujian tiap dimensi berada pada rentang 88,32%–89,88%. Agreeableness memperoleh nilai tertinggi sebesar 89,88%, sedangkan Neuroticism terendah sebesar 88,32%. Secara umum, nilai SGD pada tabel hasil lebih rendah daripada konfigurasi adaptif pada sebagian besar dimensi. SGD menggunakan laju pembelajaran yang tetap dalam konfigurasi yang diuji, sementara magnitudo gradien dapat berbeda antarparameter. Hasil yang lebih rendah pada sebagian dimensi konsisten dengan kemungkinan bahwa konfigurasi learning rate tersebut kurang sesuai untuk optimasi model ini. Pengujian learning rate dan jadwal penurunannya diperlukan untuk memisahkan pengaruh algoritma dari pengaruh pengaturan hiperparameter.

Tabel 4. Hasil Akurasi Tiap Dimensi OCEAN Optimizer RMSProp

Optimizer RMSPROP	Openness	Conscientiousness	Extraversion	Agreeableness	Neuroticism
Training	93,75%	93,42%	94,12%	93,95%	93,16%
Validation	90,61%	90,38%	90,78%	91,12%	89,71%
Testing	90,52%	90,25%	90,64%	90,85%	89,84%

Tabel 4 menyajikan hasil RMSprop. Pada subset pengujian, nilai tiap dimensi berada pada rentang 89,84%–90,85%. Agreeableness memperoleh nilai tertinggi sebesar 90,85%, sedangkan Neuroticism terendah sebesar 89,84%. Pada subset validasi, Agreeableness juga mencatat nilai tertinggi sebesar 91,12%. RMSprop menyesuaikan pembaruan parameter menggunakan rata-rata bergerak kuadrat gradien. Mekanisme ini dapat membantu ketika skala gradien berbeda antarparameter. Hasil eksperimen menunjukkan bahwa RMSprop memberi nilai validasi keseluruhan tertinggi dan nilai pengujian yang dekat dengan baseline; hasil tersebut berlaku untuk konfigurasi, pembagian data, dan metrik yang digunakan di penelitian ini.

Tabel 5. Hasil Akurasi Tiap Dimensi OCEAN Baseline

Baseline	Openness	Conscientiousness	Extraversion	Agreeableness	Neuroticism
Training	92,81%	92,52%	93,15%	92,98%	92,24%
Validation	90,50%	90,29%	90,68%	91,02%	89,61%
Testing	90,54%	90,31%	90,70%	90,92%	89,83%

Tabel 5 menyajikan hasil baseline. Pada subset pengujian, nilai tiap dimensi berkisar antara 89,83% dan 90,92%. Agreeableness memperoleh nilai tertinggi sebesar 90,92%, sedangkan Neuroticism terendah sebesar 89,83%. Nilai agregat pengujian baseline sebesar 90,46% merupakan yang tertinggi di antara konfigurasi yang dibandingkan. Baseline merupakan konfigurasi ArcFace–LSTM dengan optimizer default framework tanpa tuning hyperparameter. Nilai test 90,46% adalah hasil konfigurasi default tersebut. Karena nama optimizer default dan versinya tidak tercatat, angka ini tidak boleh ditafsirkan sebagai hasil model tanpa optimizer; pencantuman nama optimizer, learning rate default, dan versi TensorFlow pada kode diperlukan untuk replikasi.

3.3 Analisa Hasil Model

3.3.1 Analisis Komparatif Performa Optimizer

Pada validasi, RMSprop mencatat nilai agregat tertinggi (90,52%), sementara baseline tertinggi pada pengujian (90,46%). Pola ini menegaskan pentingnya membedakan hasil validasi dari hasil pengujian. RMSprop menggunakan estimasi rata-rata bergerak kuadrat gradien untuk menyesuaikan ukuran pembaruan tiap parameter, sedangkan SGD menggunakan laju global tetap pada konfigurasi yang diuji. Adam menggabungkan estimasi momen pertama dan kedua gradien dengan koreksi bias. Nilai validasi Adam sebesar 90,19% dan nilai pengujianya 90,16%, lebih rendah daripada RMSprop pada validasi dan baseline pada pengujian. Perbedaan tersebut belum cukup untuk menyimpulkan bahwa momentum Adam menyebabkan overshooting; klaim itu memerlukan bukti langsung dari lintasan loss atau eksperimen kontrol. Secara keseluruhan, data mendukung kesimpulan terbatas bahwa RMSprop merupakan konfigurasi terbaik pada validasi, sedangkan baseline menghasilkan nilai tertinggi pada pengujian. SGD memperoleh hasil agregat terendah pada kedua subset tersebut. Evaluasi tambahan dengan beberapa seed, interval ketidakpastian, dan pencarian hiperparameter yang setara dapat memperjelas kestabilan urutan kinerja ini.

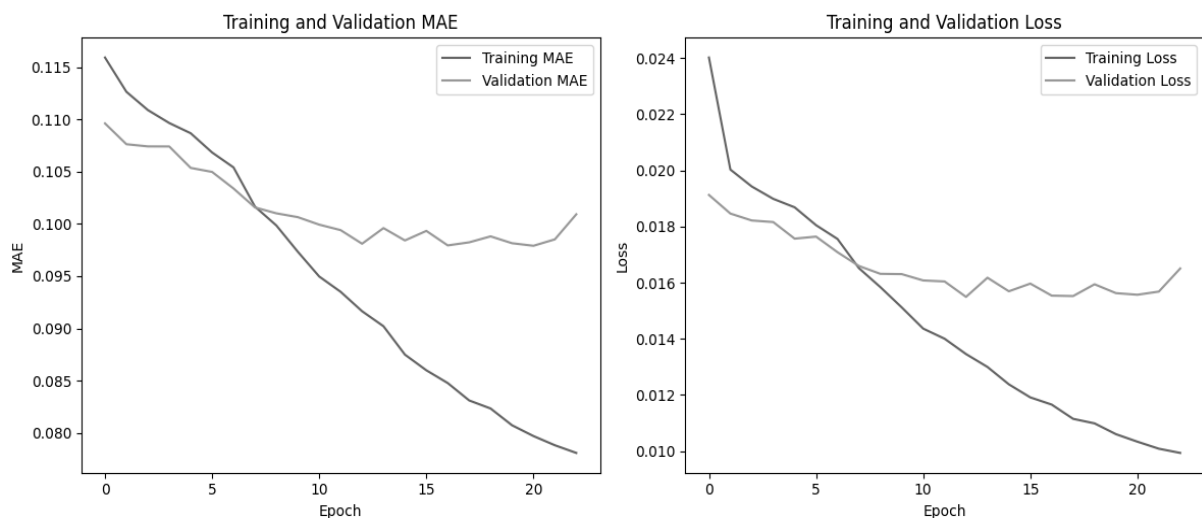
3.3.2 Transfer Learning Model ArcFace

ArcFace menggunakan Additive Angular Margin Loss untuk membentuk embedding yang diskriminatif pada tugas pengenalan wajah [8], [19]. Dalam penelitian ini, backbone ResNet34 yang telah dilatih sebelumnya dimanfaatkan sebagai ekstraktor fitur dan bobotnya dibekukan selama pelatihan model target. Hal ini mengurangi jumlah parameter backbone yang diperbarui dan menyediakan representasi fitur yang tetap bagi lapisan temporal. Embedding ArcFace

berdimensi 512 menjadi masukan bagi LSTM untuk mempelajari urutan fitur dari 10 frame. Namun, sifat diskriminatif embedding terhadap identitas wajah tidak dengan sendirinya membuktikan bahwa setiap dimensi kepribadian terwakili secara langsung. Hubungan tersebut dinilai melalui hasil regresi pada label OCEAN dan perlu ditafsirkan berdasarkan evaluasi empiris. Dengan backbone yang dibekukan, pembaruan selama pelatihan berfokus pada lapisan LSTM dan lapisan dense prediksi. Deskripsi ukuran unit dan jumlah parameter harus dibaca sesuai implementasi model yang digunakan; hasil pada naskah ini menunjukkan bahwa arsitektur tersebut dapat dilatih untuk memetakan sekuens embedding ke lima skor OCEAN.

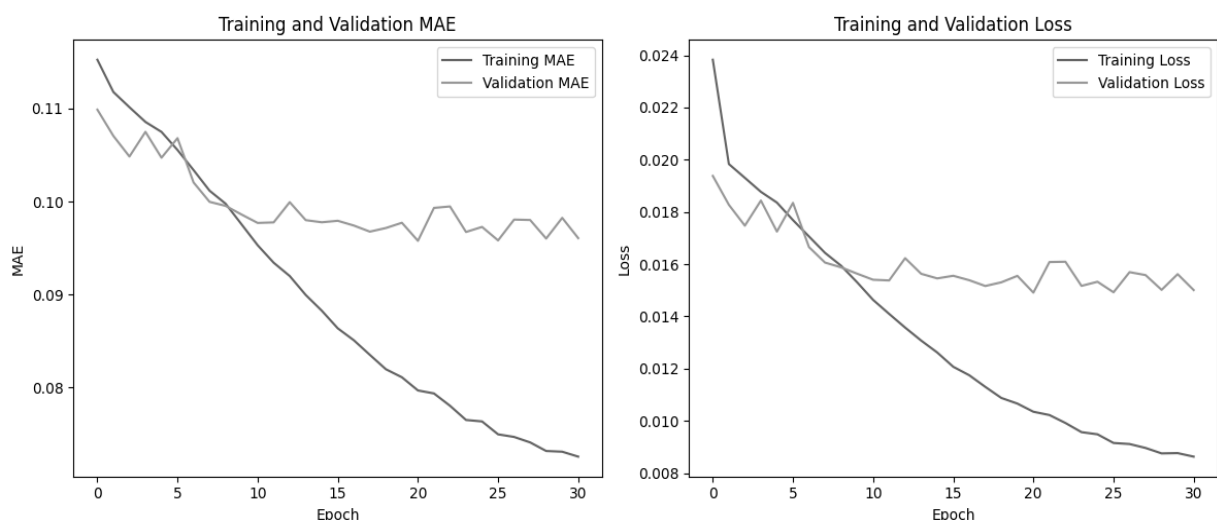
3.3.3 Dinamika Kurva Pelatihan dan Validasi

Kurva MAE dan loss pelatihan serta validasi digunakan untuk meninjau perubahan kesalahan selama epoch. MAE menunjukkan rata-rata deviasi absolut prediksi, sedangkan loss pelatihan mengikuti fungsi objektif yang digunakan untuk pembaruan parameter. Perbedaan kurva dapat memberi indikasi tentang konvergensi dan potensi overfitting, tetapi interpretasinya perlu dilihat bersama nilai akhir pada subset validasi dan pengujian. Kurva pada Gambar 6–9 memperlihatkan perubahan MAE dan loss pelatihan serta validasi untuk tiap konfigurasi.



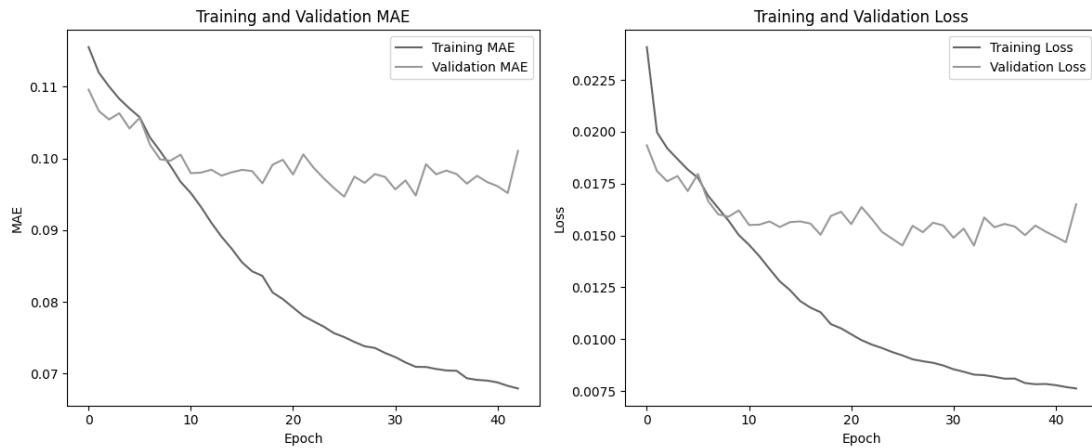
Gambar 6. Kurva Training dan Validation MAE serta Loss Optimizer Adam

Gambar 6 menunjukkan konfigurasi Adam yang berhenti pada epoch ke-22 setelah early stopping. MAE training turun dari sekitar 0,116 menjadi 0,078 dan loss training dari 0,024 menjadi 0,010. MAE validasi berada sekitar 0,098–0,101 dan loss 0,016–0,017 setelah fase awal; pola ini menggambarkan hasil pada split validasi, bukan generalisasi di luar dataset.



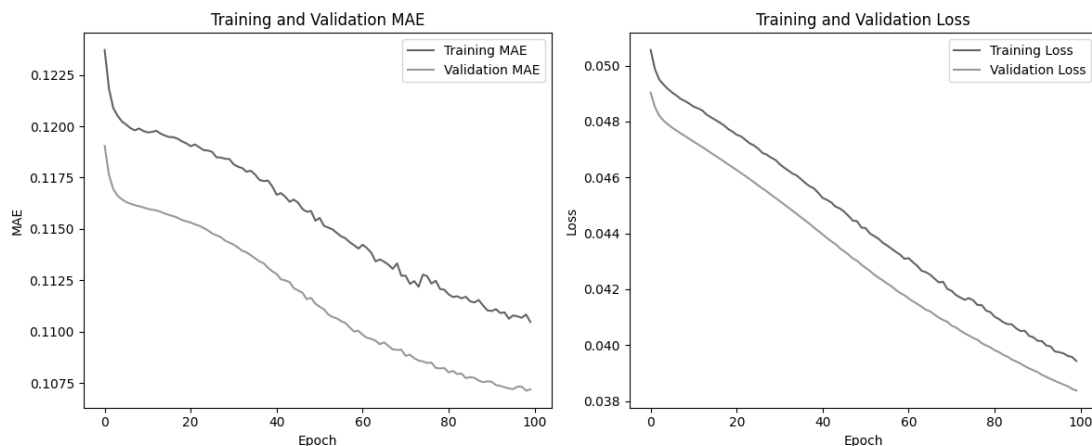
Gambar 7. Kurva Training dan Validation MAE serta Loss Baseline

Gambar 7 menunjukkan baseline berhenti pada epoch ke-30. MAE training turun dari sekitar 0,114 menjadi 0,075 dan loss training dari 0,024 menjadi 0,009. Metrik validasi berfluktuasi pada awal pelatihan, kemudian berada di sekitar MAE 0,095–0,100 dan loss 0,015–0,016. Mekanisme optimizer default perlu dicatat agar hasil baseline dapat direplikasi.



Gambar 8. Kurva Training dan Validation MAE serta Loss Optimizer RMSprop

Gambar 8 memperlihatkan RMSprop hingga epoch ke-42. MAE training turun dari sekitar 0,114 menjadi 0,068, loss training dari 0,023 menjadi 0,008, sedangkan MAE validasi berada sekitar 0,094–0,101 dan loss 0,0148–0,0165 setelah epoch ke-10. Pola ini sejalan dengan nilai validasi agregat RMSprop tertinggi pada eksperimen.



Gambar 9. Kurva Training dan Validation MAE serta Loss Optimizer SGD

Gambar 9 menunjukkan SGD berjalan hingga 100 epoch tanpa early stopping. MAE training turun dari sekitar 0,124 menjadi 0,111 dan loss dari 0,051 menjadi 0,039; MAE validasi turun dari sekitar 0,119 menjadi 0,107 dan loss dari 0,049 menjadi 0,038. Penurunan berlangsung lebih lambat pada konfigurasi yang diuji, tetapi satu kurva tidak membuktikan SGD secara umum lebih buruk.

3.4. Pembahasan Komparatif

Pada benchmark ini, frame wajah menghasilkan embedding ArcFace yang dirangkum LSTM sebelum regresi lima keluaran. Angka 1–MAE memudahkan pembacaan tetapi bukan akurasi klasifikasi; hasil terutama menggambarkan kesesuaian skor model dengan label kesan pengamat pada split ChaLearn, bukan pengukuran sifat psikologis internal. ArcFace dilatih untuk membedakan identitas wajah, sehingga embedding tidak bermakna langsung sebagai trait OCEAN [10], [19]. RMSprop memperoleh nilai validasi tertinggi (90,52%), sementara baseline memperoleh nilai test tertinggi (90,46%); selisih test kecil dan belum diuji dengan beberapa seed atau interval kepercayaan. Agreeableness tertinggi dan Neuroticism terendah pada tabel, tetapi skor agregat tidak menetapkan penyebab perbedaan tersebut. Hipotesis tentang keterlihatan ekspresi harus diuji dengan anotasi perilaku khusus.

Interpretasi hasil perlu dibatasi pada sasaran benchmark. First Impressions menyediakan penilaian pengamat atas klip singkat, sehingga label menangkap apparent personality dan dapat memuat variasi antarpemilai serta konteks perekaman. Skor prediksi tidak mengukur trait stabil secara langsung dan tidak menggantikan kuesioner tervalidasi atau asesmen psikologis [1], [20]. Karena model hanya menerima wajah, informasi bahasa, suara, dan situasi sosial tidak masuk ke prediksi; perubahan ekspresi juga dapat mencerminkan keadaan sementara. Ini membatasi pemakaian hasil sebagai kesan berbasis video, meskipun metrik agregat tampak tinggi. Pemakaian pada seleksi kerja atau keputusan individual berisiko tinggi memerlukan dasar validitas, transparansi, persetujuan, dan audit bias yang belum dievaluasi di sini [20].

Pembacaan kurva memperlihatkan bahwa nilai akhir saja belum cukup untuk memilih konfigurasi. Training MAE dan loss menurun pada seluruh optimizer, tetapi jarak antara kurva training dan validasi pada beberapa titik



mengisyaratkan bahwa kecocokan terhadap data pelatihan tidak identik dengan generalisasi. Pada RMSprop, validasi relatif berada pada rentang yang sempit setelah sekitar epoch kesepuluh, sedangkan kurva SGD menurun lebih bertahap hingga akhir pelatihan. Pola tersebut bersifat deskriptif untuk satu rangkaian eksperimen; tanpa pengulangan seed, interval ketidakpastian, dan kriteria pemilihan checkpoint yang seragam, pola kurva tidak menetapkan optimizer mana yang akan unggul pada sampel lain. Pelaporan berikutnya sebaiknya menyatakan learning rate, batch size, jumlah epoch, early stopping, seed, dan epoch checkpoint terpilih, serta menggunakan validasi hanya untuk pemilihan model sebelum evaluasi test satu kali.

Dari sisi evaluasi, transformasi $1 - \text{MAE}$ ke persen bersifat monoton urutan konfigurasi tetap sama dengan urutan MAE, tetapi istilah akurasi berpotensi disalahartikan sebagai proporsi prediksi benar. Penyebutan MAE bersama persentase pada tabel dan abstrak karena itu diperlukan agar makna angka jelas. Pelaporan MAE per trait membantu menunjukkan apakah nilai rerata menutupi dimensi yang lebih sulit, dan pelaporan korelasi atau metrik yang sensitif pada variasi antarindividu dapat menjadi pelengkap, bukan pengganti MAE. Karena keluaran berada pada skala 0–1, rincian normalisasi label, clipping keluaran, dan cara mengagregasi lima dimensi juga penting untuk reproduksi [17], [18].

Implikasi aplikatifnya terbatas pada dukungan eksploratif terhadap klip yang direkam dalam kondisi mendekati benchmark. Perubahan kamera, pencahayaan, pose, ekspresi, kelompok demografis, dan budaya dapat menggeser distribusi fitur serta persepsi label; nilai test dari satu dataset tidak menjamin kalibrasi pada lingkungan kerja, pendidikan, atau layanan publik. Studi lanjutan perlu menguji generalisasi eksternal, mendokumentasikan variasi performa antarkelompok secara bertanggung jawab, serta melibatkan penilaian manusia dan persetujuan yang sesuai. Untuk penggunaan yang berdampak pada individu, skor otomatis tidak selayaknya menjadi dasar tunggal keputusan karena kesalahan prediksi dapat membebani kelompok tertentu dan label kesan bukan ukuran kelayakan seseorang [1], [20].

4. KESIMPULAN

Studi yang dilakukan mengevaluasi prediksi lima dimensi OCEAN dari sekuens video ChaLearn First Impressions V2 menggunakan embedding wajah ArcFace dan pemodelan temporal LSTM. RMSprop memperoleh persentase akurasi regresi tertinggi pada subset validasi sebesar 90,52%, sedangkan konfigurasi yang diberi label baseline mencapai nilai tertinggi pada subset pengujian sebesar 90,46%; RMSprop memperoleh 90,42%, Adam 90,16%, dan SGD 89,27% pada pengujian. Pada RMSprop, Agreeableness mencapai 90,85% dan Neuroticism 89,84%, selaras dengan kemungkinan bahwa isyarat afiliasi lebih mudah diamati daripada variasi afek negatif yang halus, meskipun data ini tidak menguji sebab tersebut. ArcFace menyediakan embedding 512 dimensi dan LSTM merangkum urutannya untuk regresi simultan, tetapi hasil berlaku pada split dan prosedur yang dilaporkan serta tidak menjadi pengukuran kepribadian klinis. Penelitian selanjutnya perlu memastikan definisi operasional baseline dan lingkungan komputasi, mengulang eksperimen dengan beberapa seed, melaporkan MAE per dimensi dan ketidakpastian, membandingkan ArcFace dengan VGG-Face/ResNet menggunakan protokol sepadan, serta memvalidasi model lintas populasi dan modalitas sebelum penerapan praktis. Secara metodologis, kontribusi utama penelitian ini adalah pengujian pipeline visual temporal dan dokumentasi perbedaan antara hasil validasi dan test; hasil tersebut belum menunjukkan keunggulan kausal ArcFace maupun signifikansi selisih antarooptimizer. Penelitian lanjutan perlu menerbitkan konfigurasi lengkap, melakukan ablation backbone dan jumlah frame, mengevaluasi beberapa seed, serta menguji data eksternal yang beragam. Dengan batas tersebut, temuan dapat dipakai sebagai titik awal replikasi dan pengembangan alat bantu riset, bukan sebagai dasar klasifikasi atau seleksi individu.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada keluarga atas dukungan yang diberikan selama penelitian ini berlangsung. Ucapan terima kasih juga disampaikan kepada Bapak Salamun Rohman Nudin atas bimbingan dan arahnya selama proses penelitian. Penulis turut berterima kasih kepada Tim Riset dan Hanif atas bantuan dan kontribusinya dalam penyelesaian penelitian ini.

5. PERNYATAAN PENULIS

5.1 Kontribusi Penulis

Konseptualisasi: T.A.P., dan S.R.N.;

Metodologi: T.A.P., dan S.R.N.;

Perangkat Lunak: T.A.P., dan S.R.N.;

Validasi: T.A.P., dan S.R.N.;

Analisis Formal: T.A.P., dan S.R.N.;

Investigasi: T.A.P., dan S.R.N.;

Sumber Daya: T.A.P., dan S.R.N.;





Kurasi Data: T.A.P., dan S.R.N.;

Penulisan Draf Awal: T.A.P., dan S.R.N.;

Penulisan, Peninjauan, dan Penyuntingan: T.A.P., dan S.R.N.;

Visualisasi: T.A.P., dan S.R.N.;

Seluruh penulis telah membaca dan menyetujui versi manuskrip yang telah diterbitkan

5.2 Ketersediaan Data

Data yang disajikan dalam penelitian ini tersedia berdasarkan permintaan kepada penulis korespondensi.

5.3 Pendanaan

Penelitian ini didanai secara mandiri dan tidak menerima pendanaan eksternal dari instansi mana pun.

5.4 Pernyataan Dewan Peninjau Institusional

Tidak Berlaku (Not Applicable)

5.5 Pernyataan Persetujuan Berdasarkan Informasi

Tidak Berlaku (Not Applicable)

5.6 Pernyataan Konflik Kepentingan

Para penulis menyatakan bahwa mereka tidak memiliki kepentingan keuangan yang bersaing atau hubungan probadi yang diketahui yang dapat dianggap memengaruhi pekerjaan yang dilaporkan dalam artikel ini.

5.7 Penggunaan AI Generatif dan Teknologi Berbantuan AI

Para penulis menggunakan AI Generatif untuk meningkatkan kejelasan penulisan dalam artikel ini. Para penulis telah meninjau dan menyunting konten yang dihasilkan dengan bantuan AI serta bertanggung jawab penuh atas versi akhir publikasi.

REFERENCES

- [1] Suman, Chanchal; Saha, Sriparna; Gupta, Aditya; Pandey, Saurabh Kumar; Bhattacharyya, "A multi-modal personality prediction system," *Knowledge-Based Syst.*, vol. 236, 2022, doi: 10.1016/j.knosys.2021.107715.
- [2] J. F. Rauthmann, "Personality is (so much) more than just self-reported Big Five traits," vol. 38, no. 6, hal. 863–866, 2024, doi: 10.1177/08902070231221853.
- [3] Rilo, "Inferring Big-Five Personality Traits via Heterogeneous User-Post Graph Representation Learning," *R.C Pradana; M.Nasari; A.Y Zakiyyah*, hal. 1–6, 2026.
- [4] H. Zhao, Xiaoming; Liao, Yuehui; Tang, Zhiwei; Xu, Yicheng; Tao, Xin; Wang, Dandan; Wang, Guoyu; Lu, "Integrating Audio and Visual Modalities for Multimodal Personality Trait Recognition via Hybrid Deep Learning," *Front. Neurosci.*, vol. 16, 2023, doi: 10.3389/fnins.2022.1107284.
- [5] Ouarka, Ayoub; Ait Baha, Tarek; Es-Saady, Youssef; El Hajji, "A Deep Multimodal Fusion Method for Personality Traits Prediction," *Multimed. Tools Appl.*, vol. 84, no. 25, hal. 29665–29687, 2025, doi: 10.1007/s11042-024-20356-y.
- [6] G. Gan, Peter Zhuowei; Sowmya, Arcot; Mohammadi, "CLIP-based Model for Effective and Explainable Apparent Personality Perception," in *Proceedings of the 1st International Workshop on Multimodal and Responsible Affective Computing*, 2023, hal. 29–37. doi: 10.1145/3607865.3613178.
- [7] J. Wang, H. Li, W. Lok, dan S. Shan, "A single modality apparent first impression personality recognition model with temporal emotion based LSTM," *Expert Syst. Appl.*, vol. 259, no. January 2024, hal. 125114, 2025, doi: 10.1016/j.eswa.2024.125114.
- [8] Nudin, Salamun Rohman; Gernowo, Rahmat; Somantri, Maman; Wibowo, "Multi Task Classification Using Deep Learning Approaches for Big Five Personality Traits Prediction: A Review," in *Proceedings of the 11th International Conference on Information Technology, Computer and Electrical Engineering (ICITACEE 2024)*, IEEE, 2024, hal. 37–42. doi: 10.1109/ICITACEE62763.2024.10762804.
- [9] D. Minaee, Shervin; Abdolrashidi, Amirali; Su, Hang; Bennamoun, Mohammed; Zhang, "Biometrics Recognition Using Deep Learning : A Survey," *Artif. Intell. Rev.*, vol. 56, no. 8, hal. 8647–8695, 2023, doi: 10.1007/s10462-022-10237-x.
- [10] Y. Li, Zhong; Pu, Shaoyan; Lu, Jingsheng; Song, Ruibin; Zhang, Haomiao; Lu, Xuemei; Wang, "Transferring Face Recognition Techniques to Entomology: An ArcFace and ResNet Approach for Improving Dragonfly Classification," *Appl. Sci.*, vol. 15, no. 13, 2025, doi: 10.3390/app15137598.
- [11] F. Niklas, J.; Elliesen, J.; Damer, N.; Boutros, "MixQuantBio: Towards Extreme Face and Periocular Recognition Model Compression with Mixed-Precision Quantization," *Eng. Appl. Artif. Intell.*, vol. 137, 2024, doi: 10.1016/j.engappai.2024.109114.
- [12] T. B. Li, Mingyu; Bitterly, "How Perceived Lack of Benevolence Harms Trust of Artificial Intelligence Management," *J. Appl. Psychol.*, vol. 109, no. 11, hal. 1794–1816, 2024, doi: <https://doi.org/10.1037/apl0001200>.supp.
- [13] Madan, Shilpa; Park, Gayoung, "Predicting personality or prejudice ? Facial inference in the age of artificial intelligence," hal. 101815, 2024.
- [14] S. Oberoi dan Singh, "Revolutionizing Face Detection: Exploring the Potential of MTCNN Algorithm for Human Face Recognition," in *2023 Seventh International Conference on Image Information Processing (ICIIP)*, IEEE, 2023, hal. 300–306. doi: 10.1109/ICIIP61524.2023.10537691.
- [15] S. Song, Z. Shao, S. Jaiswal, L. Shen, M. Valstar, dan H. Gunes, "Learning Person-Specific Cognition From Facial Reactions





- for Automatic Personality Recognition,” *IEEE Trans. Affect. Comput.*, vol. 14, no. 4, hal. 3048–3065, 2023, doi: 10.1109/TAFFC.2022.3230672.
- [16] S. Jacques Junior, Julio C. S.; Gucluturk, Yagmur; Perez, Marc; Guclu, Umut; Andujar, Carlos; Baro, Xavier; Escalante, Hugo Jair; Guyon, Isabelle; Van Gerven, Marcel A. J.; Van Lier, Rob; Escalera, “First Impressions : A Survey on Vision-based Apparent Personality Trait Analysis,” *IEEE Trans. Affect. Comput.*, vol. 13, no. 1, hal. 75–95, 2023, doi: 10.1109/TAFFC.2019.2930058.
- [17] Y. Bounab, M. Oussalah, N. Arhab, dan S. Bekhouche, “Towards job screening and personality traits estimation from video transcriptions,” *Expert Syst. Appl.*, vol. 238, no. PD, hal. 122016, 2024, doi: 10.1016/j.eswa.2023.122016.
- [18] S. Ullah, J. Ou, Y. Xie, dan W. Tian, “Facial expression recognition (FER) survey : a vision , architectural elements , and future directions,” 2024, doi: 10.7717/peerj-cs.2024.
- [19] N. Alansari, Mohamad Yousif Abdulkareem; Hay, Oussama Abdul; Javed, Sajid; Shoufan, Abdulhadi; Zweiri, Yahya; Werghi, “GhostFaceNets: Lightweight Face Recognition Model From Cheap Operations,” *IEEE*, vol. 11, hal. 35429–35446, 2023, doi: 10.1109/ACCESS.2023.3266068.
- [20] A. Krichen, Moez; Mihoub, “Long Short-Term Memory Networks : A Comprehensive Survey,” *AI*, vol. 6, no. 9, 2025, doi: 10.3390/ai6090215.
- [21] E.-S. Tahir, Hassam; Jung, “Comparative Study on Distributed Lightweight Deep Learning Models for Road Pothole Detection,” *Sensors*, vol. 23, no. 9, 2023, doi: 10.3390/s23094347.
- [22] W. Amine, M.; Lamrani, D.; Saleh, S.; Moutaouakil, “Optimizing ResNet50 Performance Using Stochastic Gradient Descent on MRI Images for Alzheimer’s Disease Classification,” *Intell. Med.*, vol. 11, hal. 100219, 2025, doi: 10.1016/j.ibmed.2025.100219.
- [23] S. Elshamy, Reham; Abu-Elnasr, Osama; Elhoseny, Mohamed; Elmougy, “Improving the Efficiency of RMSProp Optimizer by Utilizing Nesterov in Deep Learning,” *Sci. Rep.*, vol. 13, 2023, doi: 10.1038/s41598-023-35663-x.