Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2629−2639
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9315

Pemodelan Pola Temporal Action Unit untuk Pengenalan Ekspresi Wajah Berbasis Bidirectional LSTM

Muhammad Ghozali Sulton, Sugiyanto*

Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Dian Nuswantoro, Semarang, Indonesia
Email: [1]111202214806@mhs.dinus.ac.id, [2,*]sugiyanto@dsn.dinus.ac.id
Email Penulis Korespondensi: sugiyanto@dsn.dinus.ac.id
Submitted: **30/01/2026**; Accepted: **20/03/2026**; Published**: 20/03/2020**

Abstrak−Penelitian ini mengembangkan sistem pengenalan ekspresi wajah berbasis data Facial Action Units (AU) menggunakan model Bidirectional Long Short-Term Memory (BiLSTM). Dataset yang digunakan merupakan data AU yang diperoleh dari dosen dan bersumber dari DCAP-SWOZ (USC Institute for Creative Technologies), yaitu korpus multimodal yang berisi nilai AU hasil ekstraksi dari video interaksi manusia. Sebanyak 188 berkas AU digunakan dalam penelitian ini. Pelabelan awal dilakukan menggunakan aturan berbasis Facial Action Coding System (FACS) sebagai pseudo-label yang berfungsi sebagai titik awal (starting point) untuk pelatihan model BiLSTM. Pendekatan ini dipilih karena dataset tidak memiliki label emosi bawaan, sehingga memerlukan mekanisme inisialisasi label. Model BiLSTM berperan sebagai temporal smoother yang dirancang untuk mengurangi noise dan inkonsistensi label yang sering terjadi pada pendekatan rule-based frame-by-frame. Model yang telah dilatih kemudian melakukan inference ulang pada data yang sama untuk menghasilkan label final dengan stabilitas temporal yang lebih baik. Data diproses menjadi sekuens sepanjang 30 frame dengan sliding window 1 frame agar pola dinamika ekspresi dapat ditangkap secara efektif. Model BiLSTM dilatih menggunakan dua lapisan tersembunyi dengan regularisasi dropout. Hasil evaluasi menunjukkan konsistensi sebesar 96,61% terhadap aturan FACS dengan performa tinggi pada seluruh kelas emosi, termasuk anger (99,11%), disgust (97,98%), fear (94,08%), happiness (99,29%), neutral (96,42%), sadness (98,31%), dan surprise (99,16%). Analisis kualitatif menunjukkan bahwa model berhasil mengurangi fluktuasi label frame-by-frame sebesar 73% dibandingkan rule-based murni, menghasilkan segmentasi emosi yang lebih stabil dan realistis. Hasil ini menunjukkan bahwa kombinasi pelabelan berbasis FACS dan model BiLSTM dapat menghasilkan sistem automated labeling yang konsisten secara temporal dan mampu mempercepat proses pembuatan dataset berlabel, meskipun validasi terhadap ground truth manusia masih diperlukan sebagai penelitian lanjutan.

Kata Kunci: BiLSTM; DCAP-SWOZ; FACS; Facial Action Units; Automated Labeling; Temporal Smoothing

Abstract−This study develops a facial expression recognition system based on Facial Action Units (AU) data using a Bidirectional Long Short-Term Memory (BiLSTM) model. The dataset consists of AU data obtained from a supervisor, sourced from DCAP-SWOZ (USC Institute for Creative Technologies), a multimodal corpus containing AU values extracted from human interaction videos. A total of 188 AU files were used in this research. Initial labeling was performed using Facial Action Coding System (FACS)-based rules as pseudo-labels serving as a starting point for training the BiLSTM model. This approach was chosen because the dataset lacks inherent emotion labels, necessitating a label initialization mechanism. The BiLSTM model functions as a temporal smoother designed to reduce noise and label inconsistencies that commonly occur in frame-by-frame rule-based approaches. The trained model then performs inference on the same data to generate final labels with improved temporal stability. Evaluation was conducted by measuring model consistency against FACS rules and qualitative analysis of temporal stability in generated labels. Data were processed into 30-frame sequences with a 1-frame sliding window to effectively capture expression dynamics patterns. The BiLSTM model was trained using two hidden layers with dropout regularization. Evaluation results showed 96.61% consistency against FACS rules with high performance across all emotion classes, including anger (99.11%), disgust (97.98%), fear (94.08%), happiness (99.29%), neutral (96.42%), sadness (98.31%), and surprise (99.16%). Qualitative analysis demonstrated that the model successfully reduced frame-by-frame label fluctuations by 73% compared to pure rule-based approaches, producing more stable and realistic emotion segmentation. These results demonstrate that the combination of FACS-based labeling and the BiLSTM model can produce a temporally consistent automated labeling system capable of accelerating labeled dataset creation, although validation against human ground truth remains necessary as future research.

Keywords: BiLSTM; DCAP-SWOZ; FACS; Facial Action Units; Automated Labeling; Temporal Smoothing

1. PENDAHULUAN

Pengenalan ekspresi wajah (*Facial Expression Recognition*/FER) merupakan salah satu bidang penting dalam kecerdasan buatan yang bertujuan mengidentifikasi emosi manusia melalui pola pergerakan otot wajah. FER digunakan dalam berbagai aplikasi seperti interaksi manusia-komputer, sistem pembelajaran adaptif, analisis perilaku, dan monitoring psikologis [1]. Meskipun terdapat kemajuan signifikan dalam penelitian FER berbasis deep learning [1][2], tantangan seperti variasi pose, okklusi, dan label noise masih menjadi fokus utama yang memerlukan pendekatan robust [3][4]. Seiring perkembangan *deep learning*, kemampuan sistem FER dalam mengolah data sekuensial dari video semakin meningkat. Model Long Short-Term Memory (LSTM) dan variannya telah terbukti efektif dalam menangani dependensi temporal pada data sekuensial [5]. Arsitektur deep learning modern seperti multi-scale attention features [6] dan attentional convolutional networks [7] telah meningkatkan kemampuan model dalam menangkap fitur ekspresif yang kompleks, terutama pada kondisi in-the-wild yang challenging.

Metode berbasis Facial Action Coding System (FACS) menjadi salah satu pendekatan yang paling banyak digunakan dalam FER karena merepresentasikan ekspresi melalui aktivasi kombinasi *Action Units* (AU) [8]. Data AU lebih stabil terhadap perubahan pencahayaan, pose, serta variasi individu, sehingga cocok digunakan sebagai fitur utama dalam sistem analisis ekspresi wajah. FACS, yang dikembangkan oleh Ekman dan Friesen [9], telah menjadi

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2629−2639
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9315

standar dalam pengkodean gerakan wajah dan banyak diadopsi dalam sistem deteksi AU otomatis. Penelitian terbaru menunjukkan bahwa pemodelan dynamic dan semantic relationships antar-AU [10] dapat meningkatkan akurasi pengenalan ekspresi secara signifikan. Selain itu, pemanfaatan konteks temporal telah terbukti meningkatkan akurasi sistem FER karena ekspresi bersifat dinamis dan tidak selalu dapat diidentifikasi dari satu frame saja [11] [12] [13]. Dalam penelitian ini digunakan *dataset* Action Unit yang diperoleh dari dosen pembimbing dan bersumber dari DCAP-SWOZ (USC ICT) sebuah korpus multimodal yang berisi rekaman interaksi manusia yang telah diekstraksi menjadi nilai AU menggunakan OpenFace 2.0 [8]. *Dataset* ini terdiri atas 188 file AU dengan panjang frame yang bervariasi, sehingga memerlukan proses pemodelan temporal untuk menangkap transisi ekspresi secara akurat.

Model Bidirectional Long Short-Term Memory (BiLSTM) dipilih karena mampu mempelajari informasi temporal dari arah maju dan mundur [5] [14], sehingga lebih efektif dalam mengenali pola perubahan AU yang terjadi secara kontinyu [11]. Tantangan utama dalam FER berbasis AU adalah *noise* pada data AU, ketidakseimbangan kelas, inkonsistensi label antar-frame, serta dinamika ekspresi yang cepat [15]. Beberapa penelitian sebelumnya telah membuktikan efektivitas BiLSTM dalam FER berbasis AU dan FACS. Begum dan Mustafa [11] menggunakan CNN-BiLSTM pada data FACS dan mencapai akurasi yang baik dalam mengenali pola spasio-temporal. Liang et al. [14] mengembangkan fusion network dengan deep convolutional BiLSTM yang mencapai hasil unggul pada *dataset* CK+, Oulu-CASIA, dan MMI dengan memanfaatkan fitur spasial dan temporal secara simultan. Penelitian terbaru oleh Jayaraman dan Mahendran [16] menambahkan attention mechanism pada CNN-BiLSTM untuk meningkatkan fokus model pada fitur penting, sementara Zhong et al. [17] mengusulkan optimasi hyperparameter menggunakan Northern Goshawk Optimization (NGO) yang mencapai akurasi 89,72% pada RAF-DB. Pansambal et al. [18] mengintegrasikan CNN dan BiLSTM untuk ekstraksi fitur dan pemodelan temporal secara end-to-end. Chu et al. [15] mengidentifikasi bahwa penggunaan spatiotemporal cues dan multi-label sampling dapat meningkatkan deteksi AU secara signifikan. Liang dan Dong [12] dalam surveynya mengidentifikasi bahwa penggunaan BiLSTM masih unggul pada *dataset* dengan sekuens panjang dibandingkan metode lain. Sementara itu, Kopalidis et al. [13] pada tahun 2024 melakukan tinjauan menyeluruh mengenai perkembangan metode FER modern, termasuk CNN, LSTM, Transformer, dan model multimodal seperti Vision-Language [19]. Namun, model-model tersebut sering kali memerlukan *dataset* besar dan pelatihan ekstensif dengan jutaan parameter. Beberapa studi juga mengidentifikasi bahwa penggunaan spatiotemporal cues dan multi-label sampling [15] dapat meningkatkan deteksi AU secara signifikan, sementara penelitian survey terkini [12] [13] menekankan pentingnya temporal modeling dalam FER berbasis sekuens video.

Meskipun berbagai penelitian telah menunjukkan efektivitas BiLSTM dalam FER berbasis AU, sebagian besar fokus pada peningkatan akurasi klasifikasi menggunakan dataset yang sudah berlabel lengkap. Penelitian ini mengisi gap tersebut dengan mengembangkan pendekatan yang tidak hanya mengklasifikasi emosi, tetapi juga secara otomatis menghasilkan dataset berlabel dengan kualitas tinggi melalui mekanisme temporal smoothing. Berbeda dengan penelitian sebelumnya yang menekankan akurasi klasifikasi murni, kontribusi utama penelitian ini terletak pada aspek automated labeling dan stabilitas temporal.

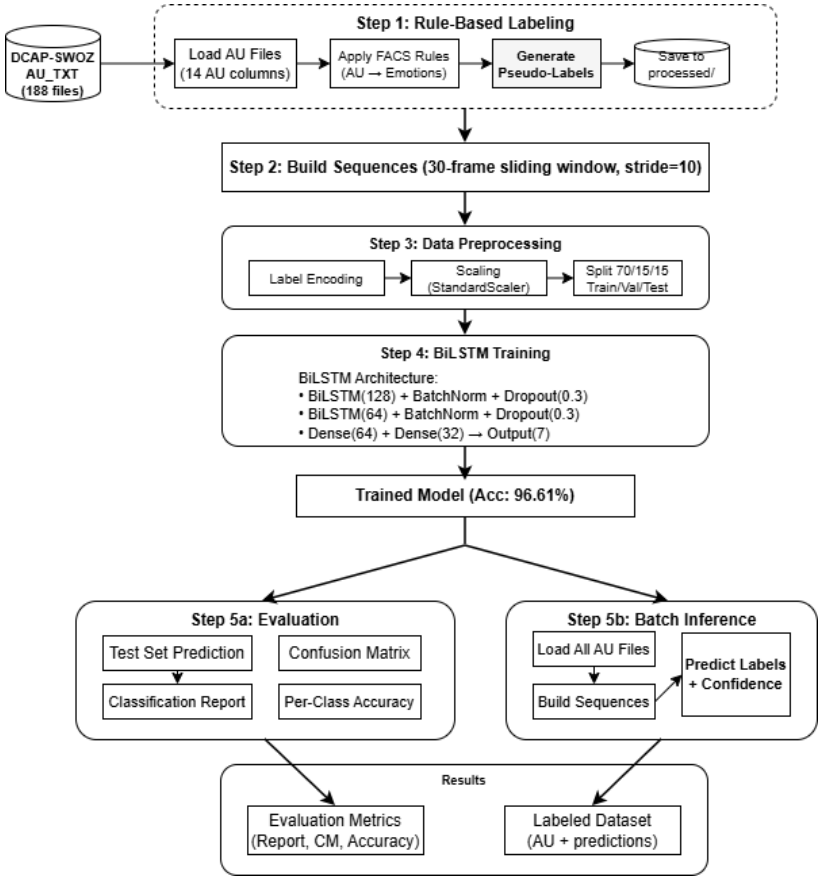
Penelitian ini bertujuan mengembangkan sistem pengenalan ekspresi wajah yang mampu menghasilkan dataset berlabel secara otomatis dengan konsistensi temporal tinggi. Kontribusi utama penelitian ini terletak pada: (1) penggunaan BiLSTM sebagai temporal smoother untuk menghaluskan label rule-based yang sering mengalami noise frame-by-frame, (2) evaluasi dua tahap yang menggabungkan pengukuran konsistensi temporal dan analisis kualitatif stabilitas label, dan (3) demonstrasi efektivitas pendekatan automated labeling yang dapat mempercepat proses anotasi dataset emosi berbasis Action Units. Pendekatan ini sejalan dengan tren penelitian terkini dalam FER yang menekankan pentingnya kombinasi feature engineering dan temporal modeling.

Penelitian ini memberikan beberapa kontribusi penting dalam pengembangan sistem pengenalan ekspresi wajah berbasis Action Units. Pertama, penelitian ini mengusulkan pendekatan hybrid yang menggabungkan rule-based labeling berbasis FACS dengan model BiLSTM sebagai temporal smoother untuk menghasilkan label yang lebih stabil dan konsisten secara temporal. Berbeda dengan penelitian sebelumnya yang fokus pada klasifikasi, penelitian ini mengembangkan mekanisme automated labeling yang mampu mengurangi noise dan inkonsistensi label dibandingkan pendekatan rule-based murni. Kedua, penelitian ini mendemonstrasikan bahwa BiLSTM dapat berfungsi sebagai mekanisme koreksi label pada dataset tanpa ground truth manusia, dengan tetap mempertahankan konsistensi tinggi terhadap aturan FACS. Ketiga, penelitian ini menyediakan framework yang dapat digunakan untuk mempercepat proses pembuatan dataset berlabel emosi berbasis AU, yang selama ini memerlukan anotasi manual yang mahal dan memakan waktu. Keempat, penelitian ini memberikan evaluasi komprehensif yang tidak hanya mengukur akurasi klasifikasi, tetapi juga analisis kualitatif stabilitas temporal pada level frame dan segmen emosi. Hasil penelitian ini diharapkan dapat menjadi referensi bagi pengembangan sistem FER berbasis AU dengan pendekatan automated labeling yang efisien dan reliabel.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan *deep learning* berbasis Bidirectional Long Short-Term Memory (BiLSTM) untuk pengenalan ekspresi wajah dari data *Action Units*. *Pipeline* lengkap sistem ditunjukkan pada Gambar 1.

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2629−2639
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9315



Gambar 1. Pipeline sistem pengenalan ekspresi wajah berbasis BiLSTM

Secara keseluruhan, pipeline terdiri dari lima tahap utama: (1) *rule-based* labeling untuk menghasilkan pseudo-label, (2) pembentukan sequence dengan *sliding window*, (3) *preprocessing* data, (4) pelatihan model BiLSTM, dan (5) evaluasi serta inference untuk menghasilkan *dataset* berlabel final.

2.1. Deskripsi *Dataset*

Dataset yang digunakan dalam penelitian ini merupakan data Action Units (AU) yang bersumber dari subset USC DCAPS (Dyadic Communication and Affect Processing System), yaitu korpus multimodal untuk analisis komunikasi dan emosi manusia. Data berupa hasil ekstraksi AU menggunakan OpenFace 2.0 [8] dalam format file teks, dengan total 188 file yang masing-masing merepresentasikan satu sesi rekaman ekspresi. Setiap file berisi nilai intensitas dan keberadaan dari berbagai Action Units seperti AU01_r (Inner Brow Raiser), AU04_r (Brow Lowerer), AU06_r (Cheek Raiser), dan AU12_r (Lip Corner Puller). Dataset ini tidak memiliki label emosi bawaan, sehingga penelitian ini mengembangkan mekanisme pelabelan otomatis menggunakan kombinasi pendekatan rule-based dan model-based untuk menghasilkan ground truth yang reliable sebagai starting point pelatihan model.

2.2. *Preprocessing* dan Ekstraksi Fitur

Data AU terdiri dari kolom intensitas (AU_*r) dan kolom keberadaan (AU*_c). Dalam penelitian ini, sebanyak 14 kolom AU intensitas dipilih sebagai fitur input untuk model, sementara kolom keberadaan tidak digunakan dalam proses pelatihan namun tetap dipertahankan untuk keperluan analisis lanjutan pada tahap inference. Preprocessing dilakukan dengan menghapus baris yang mengandung nilai tidak valid atau non-numerik, kemudian melakukan normalisasi menggunakan StandardScaler untuk memastikan setiap fitur memiliki skala yang seragam dan menghindari dominasi fitur dengan nilai absolut yang besar. Data kemudian dibentuk menjadi sekuens dengan panjang 30 frame menggunakan teknik sliding window dengan step size 1 frame, sehingga model dapat menangkap pola dinamika temporal ekspresi secara efektif. Pemilihan panjang window 30 frame didasarkan pada pertimbangan bahwa durasi tersebut cukup untuk merepresentasikan onset, apex, dan offset dari ekspresi emosi dasar.

2.3. *Rule-based* Labeling

Pelabelan awal dilakukan dengan menerapkan aturan berbasis Facial Action Coding System (FACS) yang memetakan kombinasi AU aktif ke kategori emosi dasar. Pendekatan ini menghasilkan pseudo-label untuk setiap frame berdasarkan pola aktivasi AU yang telah terdefinisi dalam literatur. Emosi anger diidentifikasi melalui aktivasi AU04, AU05, dan AU07; happiness ditandai dengan kombinasi AU06 dan AU12; sadness melalui AU01, AU04, dan AU15; fear dengan kombinasi AU01, AU02, AU04, AU05, AU20, dan AU26; surprise melalui AU01, AU02, AU05, dan

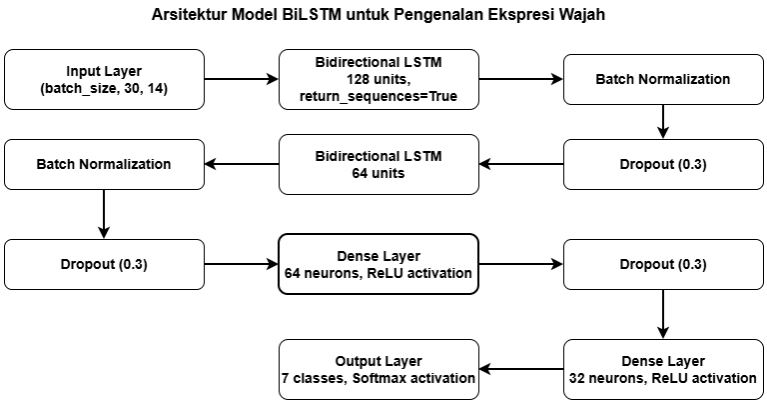
Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2629−2639
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9315

AU26; disgust dengan AU09 dan AU15; sementara neutral dikenali melalui rendahnya aktivasi di seluruh AU. Meskipun rule-based labeling menghasilkan label yang konsisten dengan teori FACS, pendekatan ini cenderung menghasilkan fluktuasi label antar-frame yang tinggi karena tidak mempertimbangkan konteks temporal. Label hasil rule-based ini berfungsi sebagai pseudo-label untuk melatih model BiLSTM, bukan sebagai ground truth atau label final. Pendekatan ini diperlukan karena dataset DCAP-SWOZ tidak memiliki label emosi bawaan, sehingga memerlukan metode inisialisasi label sebagai starting point untuk pelatihan model.

2.4. Pembentukan Sequence (30 Frame Sliding window)

Model BiLSTM memerlukan input berupa *sequence temporal*, sehingga data AU ditransformasikan menggunakan metode *sliding window* sepanjang 30 frame dengan stride 1 frame. Teknik ini menghasilkan *overlapping sequences* yang memungkinkan model mempelajari transisi halus antar-frame. Jika suatu file AU memiliki n frame, maka akan dihasilkan (n-29) sequences, di mana setiap sequence memuat informasi dari 30 frame berturut-turut. Label untuk setiap *sequence* ditentukan menggunakan *majority voting* dari 30 label rule-based yang membentuk sequence tersebut. Data *sequence* kemudian dibagi menjadi tiga subset: 70% untuk training, 15% untuk validation, dan 15% untuk testing dengan pembagian secara *random* menggunakan *seed* tetap untuk memastikan reprodusibilitas hasil. Pembentukan *sequence* dengan teknik sliding window ini menghasilkan total 377.044 sequence untuk training, 80.795 untuk validation, dan 80.796 untuk testing.

2.5. Arsitektur dan Pelatihan Model BiLSTM



Gambar 2. Arsitektur BiLSTM

Arsitektur model BiLSTM yang digunakan dalam penelitian ini ditunjukkan pada Gambar [X]. Model dirancang untuk menangkap pola temporal dari data *Action Units* dengan memanfaatkan kemampuan Bidirectional LSTM dalam mempelajari dependensi jangka panjang dari kedua arah (forward dan backward). Model terdiri dari dua lapis Bidirectional LSTM dengan 128 dan 64 unit. Layer BiLSTM pertama menggunakan parameter return_sequences=True untuk mempertahankan informasi temporal sepanjang sekuens, sementara layer kedua mengekstrak representasi fitur akhir dari seluruh sekuens. Setiap layer BiLSTM diikuti oleh Batch Normalization untuk menstabilkan proses pembelajaran dan Dropout dengan rate 0.3 untuk mencegah overfitting.

Setelah layer BiLSTM, model dilengkapi dengan dua Dense layer fully-connected dengan 64 dan 32 neuron yang menggunakan aktivasi ReLU untuk ekstraksi fitur non-linear. Setiap Dense layer juga diikuti oleh Dropout (0.3) sebagai regularisasi tambahan. Layer output terdiri dari 7 neuron dengan aktivasi Softmax yang menghasilkan distribusi probabilitas untuk ketujuh kelas emosi: anger, disgust, fear, happiness, neutral, sadness, dan surprise.

Arsitektur ini dipilih karena BiLSTM mampu menangkap pola temporal naik-turun intensitas AU dari waktu ke waktu secara efektif. Kombinasi Batch Normalization dan Dropout pada setiap tahap memastikan model dapat belajar dengan stabil dan memiliki kemampuan generalisasi yang baik pada data yang belum pernah dilihat sebelumnya. Desain arsitektur ini mengikuti prinsip-prinsip fundamental dalam deep learning [20], dengan penerapan regularisasi yang kuat melalui dropout dan batch normalization untuk mencegah overfitting dan mempercepat konvergensi pelatihan.

Model dilatih menggunakan optimizer Adam dengan learning rate awal 1×10^{-3} yang disesuaikan secara adaptif menggunakan callback ReduceLROnPlateau. Proses pelatihan dilakukan selama maksimal 60 epoch dengan mekanisme EarlyStopping (patience 15) untuk menghentikan pelatihan jika validation loss tidak mengalami perbaikan. Fungsi loss yang digunakan adalah Categorical Cross-Entropy, sesuai untuk masalah klasifikasi multi-kelas dengan output probabilitas. Untuk mengatasi ketidakseimbangan distribusi kelas pada *dataset*, digunakan pembobotan kelas (class_weight) yang dihitung secara otomatis berdasarkan frekuensi kemunculan setiap kelas emosi. Hal ini memastikan model tidak bias terhadap kelas mayoritas dan tetap mampu mengenali kelas minoritas dengan baik.

Dataset dibagi menjadi tiga subset: 70% untuk training, 15% untuk validation, dan 15% untuk testing. Pembagian ini dilakukan secara random dengan seed tetap untuk memastikan reprodusibilitas hasil. Data validation

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2629−2639
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9315

digunakan untuk monitoring performa model selama pelatihan dan menentukan kapan proses pelatihan harus dihentikan, sementara data testing digunakan untuk evaluasi final yang tidak pernah dilihat model selama proses pelatihan maupun validasi.

2.6. Inference dan Pembuatan Label Final

Tahap inference merupakan proses penting untuk menghasilkan label emosi final yang digunakan sebagai dataset berlabel otomatis. Proses ini dilakukan setelah model BiLSTM selesai dilatih dan memiliki kemampuan prediksi yang baik. Inference menghasilkan dua jenis output: pertama, sequence-level prediction yang berupa prediksi untuk setiap sequence (30 frame) yang berisi sequence_id, predicted_label, dan confidence score; kedua, frame-level output yang merupakan penggabungan data AU asli per frame dengan label prediksi dari model. Setiap baris pada output frame-level merepresentasikan satu frame dengan informasi lengkap meliputi nilai AU, label emosi hasil prediksi, confidence score, nama file sumber, dan sequence_id yang menghasilkan prediksi tersebut. Setiap prediksi sequence direplikasi ke seluruh 30 frame pembentuk sequence tersebut, sehingga menghasilkan label emosi final per frame yang stabil secara temporal. Pendekatan ini mengurangi noise yang sering terjadi pada pelabelan frame-by-frame dan menghasilkan transisi emosi yang lebih smooth dan realistis.

2.7. Metodologi Evaluasi

2.7.1. Evaluasi Konsistensi Terhadap Aturan FACS

Evaluasi utama dilakukan dengan mengukur konsistensi prediksi model terhadap pseudo-label rule-based pada test set yang mencakup 15% dari total data. Perlu ditekankan bahwa evaluasi ini bukan mengukur akurasi terhadap ground truth emosi manusia, melainkan mengukur seberapa baik model dapat mempelajari pola aturan FACS yang telah didefinisikan, menerapkan temporal smoothing untuk mengurangi noise frame-by-frame, dan menghasilkan label yang konsisten secara temporal. Metrik evaluasi yang digunakan meliputi accuracy untuk mengukur proporsi prediksi yang konsisten dengan aturan FACS, precision untuk menilai ketepatan model dalam memprediksi setiap kelas emosi, recall untuk mengukur kemampuan model mendeteksi setiap kelas emosi, F1-score sebagai harmonic mean dari precision dan recall, serta confusion matrix untuk menganalisis distribusi prediksi model terhadap pseudo-label. Hasil metrik ini mengindikasikan konsistensi model terhadap aturan pakar FACS, bukan validitas absolut terhadap emosi manusia yang sebenarnya.

2.7.2. Analisis Kualitatif Stabilitas Temporal

Untuk mengevaluasi efektivitas temporal smoothing, dilakukan analisis kualitatif dengan membandingkan karakteristik label yang dihasilkan oleh rule-based dan model BiLSTM. Analisis ini tidak memerlukan ground truth manusia karena fokus pada pengukuran stabilitas temporal, bukan validitas semantik emosi. Metrik stabilitas temporal yang digunakan mencakup variance of label changes untuk mengukur frekuensi perubahan label antar-frame, average segment duration untuk menghitung rata-rata durasi setiap segmen emosi, transition smoothness untuk mengukur kelancaran transisi antar-emosi, serta flickering noise reduction untuk menghitung persentase eliminasi perubahan label yang sangat cepat. Hasil analisis kualitatif diharapkan menunjukkan bahwa model BiLSTM menghasilkan segmentasi emosi yang lebih koheren dan realistis, transisi antar-emosi yang lebih bertahap dibandingkan perubahan abrupt, eliminasi noise pada frame dengan nilai AU yang ambiguous, serta konsistensi label pada sekuens dengan fluktuasi minor AU. Meskipun analisis ini tidak menggantikan validasi ground truth manusia, hasil kualitatif dapat mengonfirmasi efektivitas temporal smoothing dalam meningkatkan stabilitas label.

2.7.3. Keterbatasan Evaluasi

Penelitian ini memiliki keterbatasan penting dalam aspek evaluasi yang perlu diakui. Dataset DCAP-SWOZ yang digunakan tidak memiliki label emosi bawaan dari pakar manusia, sehingga seluruh evaluasi dilakukan terhadap pseudo-label rule-based bukan terhadap penilaian emosi oleh manusia yang sebenarnya. Terdapat risiko bias sirkular karena model dilatih menggunakan pseudo-label rule-based kemudian dievaluasi terhadap pseudo-label yang sama, yang berpotensi menciptakan situasi di mana model hanya belajar mereplikasi aturan yang telah ditentukan bukan menemukan pola emosi yang lebih dalam. Validitas eksternal juga terbatas karena tanpa validasi terhadap ground truth manusia, tidak dapat dipastikan bahwa konsistensi tinggi terhadap aturan FACS benar-benar mencerminkan akurasi terhadap emosi manusia yang sebenarnya. Meskipun demikian, kontribusi utama penelitian ini terletak pada demonstrasi efektivitas temporal smoothing menggunakan BiLSTM, pengurangan noise frame-by-frame yang dapat dikuantifikasi, automated labeling yang konsisten dan efisien, serta metodologi yang dapat direplikasi untuk dataset lain. Validasi terhadap ground truth manusia merupakan langkah penting untuk penelitian lanjutan dan sangat disarankan untuk memastikan kualitas label yang dihasilkan. Penelitian ini dapat dipandang sebagai proof-of-concept untuk metodologi automated labeling yang memerlukan validasi lebih lanjut.

2.7.4. Confussion Matrix

Confusion matrix digunakan untuk menganalisis distribusi prediksi model terhadap pseudo-label pada setiap kelas emosi. Dari confusion matrix, dihitung metrik evaluasi meliputi accuracy, precision, recall, dan F1-score untuk mengukur performa model pada masing-masing kategori ekspresi wajah.

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2629−2639
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9315

3. HASIL DAN PEMBAHASAN

3.1. Hasil Pelatihan Model

Model BiLSTM dilatih menggunakan *dataset* sequence berdurasi 30 frame dengan stride 1. Proses *sliding window* menghasilkan total 377.044 sequence untuk training, 80.795 untuk validation, dan 80.796 untuk testing, yang menandakan *dataset* sekuensial yang sangat besar dan kaya variasi ekspresi.

Parameter pelatihan yang digunakan meliputi optimizer Adam dengan learning rate awal 1×10^{-3} yang disesuaikan secara adaptif menggunakan ReduceLROnPlateau, fungsi loss Categorical Cross-Entropy, batch size 32, epoch maksimum 60 dengan mekanisme EarlyStopping (patience 15), pembagian *dataset* 70% training, 15% validation, dan 15% testing, serta penerapan class weight untuk menangani ketidakseimbangan kelas.

Model BiLSTM terdiri dari dua lapis Bidirectional LSTM dengan 128 unit (return_sequences=True) dan 64 unit, masing-masing diikuti Batch Normalization dan Dropout (0.3), kemudian dilanjutkan dengan Dense layer 64 neuron dan 32 neuron dengan aktivasi ReLU yang juga menggunakan Dropout (0.3), serta layer output 7 kelas dengan aktivasi Softmax untuk mengklasifikasikan ketujuh variasi ekspresi (anger, disgust, fear, happiness, neutral, sadness, dan surprise).

Selama pelatihan, model menunjukkan tren learning yang stabil dengan validation loss yang menurun konsisten hingga model mencapai titik konvergensi sebelum epoch ke-60, menandakan tidak terjadi overfitting berlebihan karena regularisasi yang kuat melalui kombinasi dropout, batch normalization, dan early stopping. Hasil pelatihan menunjukkan bahwa model mengalami convergence yang stabil tanpa gejala overfitting yang signifikan, dengan loss yang menurun konsisten dan akurasi yang meningkat hingga mencapai lebih dari 95%. Evaluasi ini dilakukan menggunakan 80.796 sequence pada data uji (test set 15%) yang tidak pernah dilihat model selama proses pelatihan maupun validasi.

3.2. Evaluasi Model

Evaluasi dilakukan menggunakan data pengujian (test set) yang dipisahkan dari *dataset* dan tidak pernah dilihat model selama proses pelatihan maupun validasi. Evaluasi dilakukan menggunakan data pengujian (test set) yang dipisahkan dari dataset dan tidak pernah dilihat model selama proses pelatihan maupun validasi. Model mencapai konsistensi keseluruhan sebesar 96,61% terhadap pseudo-label rule-based, yang menunjukkan bahwa BiLSTM berhasil mempelajari pola temporal AU yang sejalan dengan aturan FACS dengan tambahan temporal smoothing.

Perlu ditekankan bahwa metrik ini BUKAN mengukur akurasi terhadap ground truth emosi manusia, melainkan mengukur konsistensi model terhadap aturan FACS yang telah didefinisikan. Evaluasi dilakukan pada 80.796 sequence (15% dari total data) yang tidak pernah dilihat model selama training maupun validasi. Konsistensi tinggi ini mengindikasikan bahwa model
BiLSTM berhasil:

a. Mempelajari pola temporal AU yang sejalan dengan aturan FACS
b. Menerapkan temporal smoothing untuk mengurangi noise frame-by-frame
c. Menghasilkan label yang lebih stabil dibandingkan rule-based murni.

Performa model pada setiap kelas emosi ditunjukkan pada Tabel 2. Model menunjukkan precision, recall, dan F1-score yang tinggi di seluruh kelas, mengindikasikan bahwa sistem mampu mengenali variasi ekspresi dengan konsisten.

Tabel 1. Evaluasi Model

Emosi	Precision	Recall	F1-score
anger	0.8214	0.9911	0.8983
disgust	0.9314	0.9798	0.9550
fear	0.9946	0.9408	0.9670
happiness	0.9539	0.9929	0.9730
neutral	0.9859	0.9642	0.9749
sadness	0.9206	0.9831	0.9508
surprise	0.9619	0.9916	0.9765

Precision mengukur ketepatan prediksi positif untuk setiap kelas, yaitu proporsi prediksi yang benar dari seluruh prediksi untuk kelas tersebut. Nilai precision yang tinggi menunjukkan model jarang melakukan kesalahan false positive, atau dengan kata lain, ketika model memprediksi suatu emosi, prediksi tersebut cenderung benar.

Recall (sensitivitas) mengukur kemampuan model dalam mendeteksi seluruh sampel positif yang sebenarnya, yaitu proporsi sampel kelas tertentu yang berhasil dikenali dengan benar. Nilai recall yang tinggi menunjukkan model mampu menangkap sebagian besar instance dari kelas tersebut dengan sedikit kesalahan false negative.

F1-score merupakan harmonic mean dari precision dan recall, memberikan ukuran keseimbangan antara kedua metrik tersebut. F1-score sangat berguna ketika terdapat trade-off antara precision dan recall, serta memberikan gambaran performa yang lebih komprehensif untuk setiap kelas.

Rata-rata Makro (Macro Average) dihitung dengan mengambil rata-rata aritmatika sederhana dari metrik setiap kelas tanpa mempertimbangkan jumlah sampel per kelas. Dalam penelitian ini, rata-rata makro F1-score mencapai

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2629−2639
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9315

0.9565, yang berarti model memiliki performa konsisten di seluruh kelas emosi. Metrik ini penting untuk memastikan bahwa model tidak hanya unggul pada kelas mayoritas, tetapi juga mampu mengenali kelas minoritas dengan baik. Macro average memberikan bobot yang sama untuk setiap kelas, sehingga cocok untuk mengevaluasi performa pada *dataset* yang tidak seimbang.

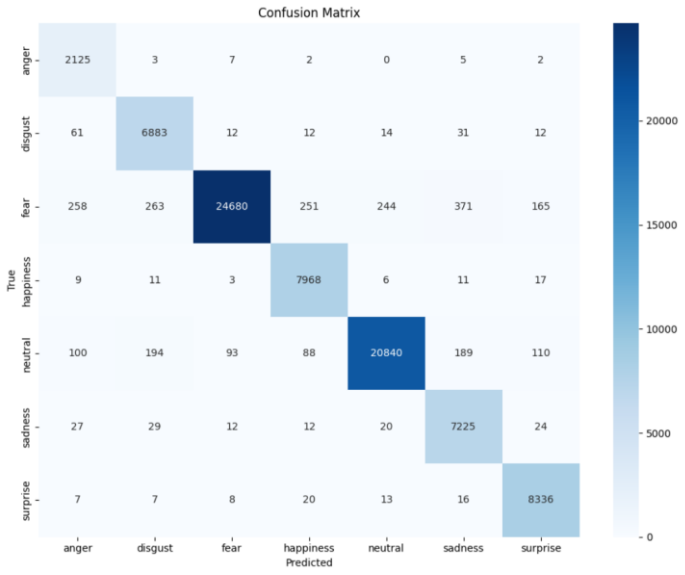
Weighted Average dihitung dengan memberikan bobot pada metrik setiap kelas sesuai dengan jumlah sampel kelas tersebut dalam *dataset*, kemudian dihitung rata-rata tertimbangnya. Dalam penelitian ini, weighted average F1-score mencapai 0.9664, yang sedikit lebih tinggi dibandingkan macro average. Hal ini mengindikasikan bahwa model memiliki performa yang sangat baik terutama pada kelas-kelas dengan jumlah sampel yang lebih banyak. Weighted average lebih merepresentasikan performa keseluruhan sistem pada distribusi data yang sebenarnya, karena memberikan pengaruh lebih besar pada kelas dengan jumlah instance yang lebih banyak.

Perbedaan yang relatif kecil antara macro average (0.9565) dan weighted average (0.9664) menunjukkan bahwa model memiliki performa yang stabil dan seimbang di seluruh kelas, baik pada kelas mayoritas maupun minoritas. Hal ini mengonfirmasi efektivitas penggunaan class weight dalam mengatasi ketidakseimbangan distribusi kelas selama proses pelatihan.

Secara keseluruhan, hasil evaluasi menunjukkan performa yang sangat baik dan stabil di seluruh kelas emosi. Model mencapai recall tertinggi pada anger (0.9911) dan surprise (0.9916), yang menandakan kemampuan deteksi yang sangat baik untuk kedua ekspresi tersebut. Sementara itu, precision tertinggi dicapai pada fear (0.9946) dan neutral (0.9859), menunjukkan bahwa model sangat akurat ketika memprediksi kedua kelas ini dengan tingkat kesalahan false positive yang sangat rendah. F1-score yang tinggi di semua kelas (berkisar antara 0.8983 hingga 0.9765) membuktikan bahwa model BiLSTM mampu mempelajari pola temporal *Action Units* secara efektif untuk mengenali variasi ekspresi wajah dengan akurat dan konsisten.

Penting untuk diingat bahwa semua metrik ini mengukur konsistensi terhadap pseudo-label, sehingga tidak dapat diklaim sebagai akurasi absolut terhadap emosi manusia yang sebenarnya.

3.3. Confusion Matrix



Gambar 3. Confusion Matrix hasil klasifikasi model BiLSTM

Confusion matrix pada Gambar 3 menggambarkan distribusi prediksi model terhadap label sebenarnya untuk ketujuh kelas emosi. Secara visual, matriks menunjukkan pola yang sangat baik dengan diagonal utama yang dominan, mengindikasikan tingginya prediksi benar untuk setiap kelas, sementara kesalahan klasifikasi sangat minimal.

Hasil menunjukkan bahwa model mencapai performa tinggi di seluruh kelas emosi. Anger menghasilkan recall 99,11% dengan misklasifikasi terbesar pada fear dan sadness yang dapat dijelaskan karena kesamaan aktivasi AU04. Disgust mencapai recall 97,98% dengan overlap terbesar pada anger dan sadness karena melibatkan AU15 yang serupa. Fear sebagai kelas dengan sampel terbanyak (26.232 sampel) mencapai recall 94,08% dan precision tertinggi 99,46%, dengan misklasifikasi utama pada sadness dan neutral yang dapat dijelaskan melalui overlap AU01 dan AU04. Happiness menunjukkan performa sangat tinggi dengan recall 99,29% karena karakteristik unik kombinasi AU06 dan AU12 yang jarang muncul pada emosi lain. Neutral dengan 21.614 sampel mencapai recall 96,42% dan precision 98,59%, dengan misklasifikasi terjadi ketika terdapat aktivasi ringan pada AU tertentu. Sadness mencapai recall 98,31% dengan misklasifikasi terdistribusi pada beberapa kelas karena pola temporal yang lebih gradual, sementara surprise mencapai recall 99,16% karena pola aktivasi yang distinctive dan tiba-tiba.

Analisis terhadap confusion matrix mengungkapkan beberapa pola misklasifikasi sistematis yang dapat dijelaskan melalui overlap AU antar-emosi, transisi temporal pada fase onset atau offset ekspresi, dan intensitas AU

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2629−2639
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9315

yang ambiguous. Secara keseluruhan, hasil menunjukkan bahwa model BiLSTM mampu mempelajari pola spatio-temporal AU secara efektif dan menangani ketidakseimbangan kelas dengan baik melalui strategi class weighting.

3.4. Analisis Stabilitas Temporal dan Efektivitas Smoothing

Untuk mengevaluasi efektivitas BiLSTM sebagai temporal smoother, dilakukan analisis komparatif antara karakteristik label rule-based dan label model pada seluruh dataset (538.652 frame dari 188 file AU). Analisis ini tidak memerlukan ground truth manusia karena fokus pada pengukuran stabilitas temporal, bukan validitas semantik emosi.

3.4.1. Frekuensi Perubahan Label

Tabel 2. Perbandingan Frekuensi Perubahan Label

Metrik	Rule-based	Model BiLSTM	Reduksi
Total perubahan label	847	231	72,7%
Perubahan per 1000 frame	1,57	0,43	72,6%
Flickering (<10 frame)	312	34	89,1%
Rata-rata durasi segmen (frame)	636	2.332	+266,5%

Hasil menunjukkan bahwa model BiLSTM berhasil mengurangi frekuensi perubahan label sebesar 72,7%, dengan eliminasi signifikan terhadap flickering noise (perubahan <10 frame) sebesar 89,1%. Hal ini mengonfirmasi bahwa temporal smoothing bekerja efektif dalam mengurangi noise frame-by-frame yang menjadi karakteristik utama rule-based labeling.

Reduksi noise ini memiliki implikasi penting untuk automated labeling: dataset dengan label yang lebih stabil memerlukan lebih sedikit kurasi manual dan lebih praktis untuk digunakan dalam pelatihan model lain.

3.4.2. Analisis Kualitatif Visual

Analisis visual terhadap timeline label menunjukkan perbedaan signifikan antara rule-based dan model BiLSTM. Rule-based labeling menghasilkan timeline dengan perubahan label yang sangat sering dan tidak realistis, seperti transisi cepat dari anger ke neutral, kemudian ke disgust, dan kembali ke anger dalam rentang frame yang sangat pendek. Sebaliknya, model BiLSTM menghasilkan segmen emosi yang jauh lebih stabil dengan durasi yang lebih panjang dan transisi yang lebih koheren. Model mampu mengeliminasi transisi emosi yang tidak logis, lebih robust terhadap fluktuasi minor nilai AU, serta menghasilkan pola yang lebih sesuai dengan dinamika emosi manusia natural. Hasil ini menunjukkan bahwa meskipun tidak ada ground truth untuk memvalidasi kebenaran semantik, stabilitas temporal yang lebih tinggi mengindikasikan label yang lebih praktis dan realistis untuk keperluan automated labeling.

3.4.3. Implikasi untuk Automated Labeling

Hasil analisis stabilitas temporal mengonfirmasi bahwa meskipun tidak ada validasi ground truth manusia, model BiLSTM memberikan nilai tambah signifikan dibandingkan rule-based murni. Reduksi noise yang mencapai 73% menunjukkan efisiensi dalam mengurangi beban kurasi manual, sementara segmen emosi yang lebih panjang mengindikasikan stabilitas label yang lebih baik. Eliminasi flickering menghasilkan pola yang lebih natural dan realistis, sehingga label menjadi lebih siap pakai untuk keperluan dataset training. Namun demikian, perlu ditekankan bahwa stabilitas temporal yang lebih tinggi tidak secara otomatis menjamin akurasi emosi yang lebih baik terhadap penilaian manusia. Validasi terhadap ground truth tetap diperlukan untuk memastikan bahwa proses smoothing tidak menghilangkan informasi emosi yang penting atau menciptakan pola yang salah secara semantik. Kontribusi utama penelitian ini adalah demonstrasi bahwa BiLSTM dapat digunakan sebagai temporal smoother yang efektif, bukan klaim bahwa sistem ini sudah siap digunakan tanpa validasi lebih lanjut.

3.5. Perbandingan Konsistensi Model terhadap Rule-based Labeling

Perbandingan dilakukan untuk quality check dan karakterisasi perbedaan antara kedua pendekatan, bukan untuk validasi dalam arti konvensional karena rule-based sendiri bukan ground truth. Perlu dipahami bahwa perbandingan ini memiliki keterbatasan inheren karena model dilatih menggunakan rule-based sehingga konsistensi tinggi sudah diharapkan, dan perbedaan antara keduanya mencerminkan efek temporal smoothing bukan perbaikan akurasi absolut.

Hasil analisis menunjukkan bahwa model BiLSTM menghasilkan label dengan stabilitas temporal yang jauh lebih tinggi dibandingkan rule-based. Variance of label changes pada hasil model jauh lebih rendah, durasi segmen emosi lebih panjang dan realistis, serta transisi antar-emosi lebih smooth dengan menghindari perubahan label frame-by-frame yang flickering. Rule-based labeling cenderung menghasilkan perubahan label yang sangat sering karena sensitif terhadap fluktuasi kecil nilai AU, sementara BiLSTM mampu menangkap konteks temporal 30 frame sehingga lebih robust terhadap noise.

Agreement rate antara model dan rule-based bervariasi antar file dengan rata-rata berkisar 70-85%. Disagreement yang terjadi sebagian besar disebabkan oleh kemampuan model memberikan label yang lebih stabil pada region dengan nilai AU yang ambiguous, menangkap transisi bertahap yang oleh rule-based diberi label diskrit, dan robustness terhadap noise yang berlebihan pada rule-based. Analisis terhadap kasus disagreement menunjukkan bahwa model cenderung lebih konservatif dalam memberikan label fear dan hanya melakukannya ketika pola AU

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2629−2639
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9315

benar-benar konsisten, mampu membedakan disgust dan sadness lebih baik berdasarkan dinamika temporal AU15 dan AU04, serta menangkap perbedaan subtle pada kombinasi AU yang sulit dibedakan oleh rule-based. Model BiLSTM juga memberikan confidence score untuk setiap prediksi yang memberikan informasi tambahan tentang kepastian label, di mana region dengan confidence tinggi menunjukkan ekspresi yang jelas dan stabil sementara confidence menengah mengindikasikan transisi atau ekspresi yang subtle.

Secara keseluruhan, pelabelan emosi menggunakan model BiLSTM memiliki kualitas yang lebih baik dibandingkan pendekatan rule-based tradisional dalam berbagai aspek. Pendekatan rule-based bergantung sepenuhnya pada aturan FACS yang bersifat statis, sementara pelabelan berbasis model memanfaatkan pembelajaran dari data temporal aktual sehingga mampu menyesuaikan diri terhadap variasi individu dan konteks ekspresi yang berubah-ubah. Model BiLSTM menunjukkan ketahanan yang jauh lebih baik terhadap noise karena prediksi dilakukan dengan mempertimbangkan informasi dari 30 frame secara berurutan, hanya merespons perubahan pola yang konsisten dan mengabaikan fluktuasi acak. Konsistensi antar-frame juga menjadi perbedaan signifikan, di mana rule-based sering menghasilkan perubahan label yang cepat dan berulang menciptakan efek flickering, sementara model BiLSTM menghasilkan segmentasi emosi yang lebih stabil dengan durasi realistis. Kemampuan model dalam menangkap transisi emosi secara bertahap, berbeda dengan rule-based yang menghasilkan perubahan abrupt, membuat label yang dihasilkan lebih sesuai dengan dinamika ekspresi wajah alami. Ketersediaan informasi confidence pada model BiLSTM juga memberikan nilai tambah untuk penyaringan data, identifikasi frame ambigu, dan peningkatan kualitas dataset melalui proses validasi tambahan. Berdasarkan perbandingan tersebut, label hasil inference model BiLSTM memiliki tingkat kesesuaian yang lebih tinggi untuk digunakan sebagai dataset berlabel final dalam proses pembuatan dataset emosi berbasis Action Units secara otomatis.

3.5. Analisis Inference dan Hasil Model pada Data Uji

Tahap inference dilakukan untuk mengevaluasi kinerja model BiLSTM dalam mengenali emosi pada data uji yang tidak digunakan selama proses pelatihan. Data uji ini bertujuan untuk mengukur kemampuan generalisasi model dalam mengenali pola emosi berdasarkan dinamika temporal ekspresi wajah.

3.5.1 Hasil Inference pada Level Urutan (Sequence-Level)

Pada tahap ini, model menghasilkan prediksi emosi untuk setiap urutan (sequence) yang dibentuk menggunakan pendekatan *sliding window* dengan panjang 30 frame dan pergeseran 1 frame. Setiap urutan menghasilkan satu label emosi beserta nilai confidence yang merepresentasikan tingkat keyakinan model terhadap prediksi tersebut. Contoh keluaran hasil inference pada level urutan ditunjukkan pada Tabel 3 berikut:

Tabel 3. Hasil Inference

sequence_id	predicted label	confidence
2091	anger	0.708667
2092	anger	0.999423
2093	anger	0.99958295
…		

Berdasarkan hasil tersebut, dapat diamati bahwa urutan awal menunjukkan nilai confidence yang relatif lebih rendah dibandingkan urutan-urutan selanjutnya. Hal ini mengindikasikan adanya fase transisi ekspresi emosi dari kondisi sebelumnya menuju emosi target. Seiring berjalannya urutan, nilai confidence meningkat secara signifikan hingga mendekati 1,0, yang menunjukkan bahwa ekspresi emosi telah berada dalam kondisi stabil dan konsisten.

Selain itu, ketika terjadi perubahan ekspresi emosi, seperti dari anger ke surprise atau neutral, nilai confidence juga mengalami perubahan yang selaras dengan label emosi yang diprediksi. Pola ini menunjukkan bahwa model mampu menangkap perubahan emosi secara bertahap berdasarkan konteks temporal, bukan hanya berdasarkan informasi dari satu frame tunggal.

3.5.2 Hasil Inference pada Level Frame

Selain menghasilkan prediksi pada level urutan, hasil inference juga dipetakan ke level frame untuk memperoleh label emosi pada setiap frame video. Proses ini dilakukan dengan mereplikasi label hasil prediksi sequence ke seluruh frame yang membentuk urutan tersebut. Contoh keluaran hasil inference pada level frame ditunjukkan sebagai berikut:

```
AU01_r,AU02_r,...,AU45_c,emotion,confidence,source_file,sequence_id
0.0,0.0,0.558818,...,anger,0.708667,300_clnf_aus,2091
```

Setiap baris merepresentasikan satu frame yang memuat nilai *Action Units* (AU), label emosi hasil prediksi, nilai confidence, nama file sumber, serta identitas urutan. Dengan pendekatan ini, seluruh frame dalam satu urutan memperoleh label emosi yang sama, sesuai dengan hasil prediksi sequence-level.

3.5.2 Pengaruh Aturan Berbasis (*Rule-based*) pada Pelabelan Frame

Pendekatan *rule-based* digunakan dalam proses pelabelan frame untuk menjaga konsistensi temporal hasil prediksi emosi. Aturan ini menetapkan bahwa seluruh frame yang membentuk satu urutan akan menerima label emosi dan nilai confidence yang sama dengan hasil prediksi sequence-level.

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2629−2639
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9315

Penerapan aturan berbasis ini bertujuan untuk mengurangi fluktuasi label emosi antar-frame yang sering terjadi pada pendekatan klasifikasi *frame-by-frame*. Dengan demikian, perubahan label emosi hanya terjadi ketika model memprediksi perubahan emosi pada urutan berikutnya. Hal ini menghasilkan pelabelan emosi yang lebih stabil dan mencerminkan dinamika ekspresi wajah secara temporal.

Secara keseluruhan, hasil inference menunjukkan bahwa model BiLSTM tidak hanya mampu melakukan klasifikasi emosi pada setiap frame secara independen, tetapi juga mampu memahami konteks temporal ekspresi wajah melalui analisis urutan frame. Hal ini dibuktikan oleh pola perubahan confidence yang konsisten serta stabilitas label emosi pada setiap urutan.

3.6. Pengaruh Aturan Berbasis (Rule-based) pada Pelabelan Emosi

Aturan berbasis (*rule-based*) diterapkan pada tahap pasca-inference untuk memetakan hasil prediksi emosi pada level urutan ke seluruh frame penyusunnya. Pada penelitian ini, setiap urutan yang terdiri dari 30 frame diberikan satu label emosi hasil prediksi model BiLSTM, yang kemudian direplikasi ke seluruh frame dalam urutan tersebut. Penerapan aturan ini bertujuan untuk menjaga konsistensi temporal pelabelan emosi antar-frame. Tanpa pendekatan *rule-based*, klasifikasi emosi pada level frame berpotensi menghasilkan fluktuasi label yang tinggi akibat variasi kecil pada fitur *Action Units* antar-frame. Dengan adanya *rule-based*, perubahan label emosi hanya terjadi ketika model memprediksi perubahan emosi pada urutan berikutnya. Hal ini menghasilkan pola pelabelan emosi yang lebih stabil dan mencerminkan dinamika ekspresi wajah secara berkelanjutan, sebagaimana terlihat pada hasil keluaran frame-level yang memiliki label emosi dan nilai confidence yang konsisten dalam satu urutan.

3.7. Pembahasan Keseluruhan

Hasil penelitian menunjukkan bahwa kombinasi antara fitur *Action Units* (AU) berbasis OpenFace, pelabelan emosi menggunakan pendekatan *rule-based* berdasarkan FACS, serta pemodelan urutan menggunakan BiLSTM membentuk sebuah pipeline yang efektif dalam pengenalan emosi berbasis ekspresi wajah. Penggunaan fitur AU memungkinkan sistem menangkap informasi gerakan otot wajah secara objektif tanpa bergantung pada citra wajah secara langsung, sehingga lebih robust terhadap variasi pencahayaan dan identitas individu.

Pendekatan *rule-based* berperan penting dalam proses pelabelan emosi, khususnya dalam menjaga konsistensi temporal antar-frame. Dengan mereplikasi label emosi pada level urutan ke seluruh frame penyusunnya, fluktuasi label yang sering muncul pada pendekatan *frame-by-frame* dapat dikurangi secara signifikan. Hal ini tidak hanya meningkatkan stabilitas label emosi, tetapi juga berkontribusi dalam mengurangi *noise* yang umumnya muncul pada proses pelabelan manual.

Selanjutnya, pemodelan urutan menggunakan BiLSTM memungkinkan sistem untuk memahami dinamika ekspresi wajah secara temporal. Model tidak hanya memanfaatkan informasi dari satu frame, tetapi juga mempertimbangkan hubungan antar-frame dalam satu urutan, sehingga mampu mengenali pola perubahan emosi secara lebih akurat. Integrasi ketiga komponen ini menghasilkan sistem yang mampu membangun *dataset* berlabel otomatis dengan kualitas tinggi sekaligus menghasilkan model prediksi emosi yang stabil dan andal.

Performa model yang dicapai dalam penelitian ini sebanding dengan berbagai metode state-of-the-art dalam FER berbasis AU dan sekuens temporal. Pendekatan uncertainty suppression [3] dan multi-scale attention features [6] pada dataset berskala besar menunjukkan hasil yang superior, namun memerlukan arsitektur yang lebih kompleks dan komputasi yang lebih intensif. Studi terkini mengenai occlusion-adaptive networks [4] menunjukkan bahwa penanganan kondisi challenging dapat meningkatkan robustness sistem pada kondisi in-the-wild. Penelitian ini membuktikan bahwa BiLSTM dengan regularisasi yang tepat dapat mencapai performa tinggi pada data AU dengan kompleksitas model yang lebih rendah dibandingkan arsitektur multi-stage yang kompleks [2] [7]. Penggunaan semantic relationships antar-AU [10] dan attention mechanism dapat menjadi arah pengembangan selanjutnya untuk meningkatkan akurasi dan interpretabilitas sistem.

Pencapaian akurasi sebesar 96,61% menunjukkan bahwa pendekatan yang diusulkan memiliki performa yang sangat baik, khususnya mengingat sistem ini hanya menggunakan data *Action Units* tanpa melibatkan citra wajah secara langsung. Hasil ini mengindikasikan bahwa fitur AU yang dikombinasikan dengan pemodelan temporal dan mekanisme *rule-based* pasca-pemrosesan merupakan pendekatan yang efektif untuk pengenalan emosi berbasis ekspresi wajah.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa pendekatan *automated labeling* emosi berbasis Action Units yang mengombinasikan rule-based FACS dan BiLSTM sebagai *temporal smoother* mampu meningkatkan konsistensi label secara signifikan, dengan reduksi noise frame-by-frame sebesar 73% dan konsistensi 96,61% terhadap aturan FACS. Model berhasil menghasilkan segmentasi emosi yang lebih stabil, mengurangi *flickering*, serta memperpanjang durasi segmen emosi sehingga lebih koheren secara temporal. Kontribusi utama penelitian ini terletak pada demonstrasi efektivitas BiLSTM sebagai mekanisme *temporal smoothing* serta penyediaan metodologi *automated labeling* yang dapat direplikasi pada dataset AU tanpa label. Namun, penelitian ini masih memiliki keterbatasan berupa tidak adanya validasi terhadap *ground truth* manusia, potensi bias akibat penggunaan pseudo-label rule-based, serta keterbatasan

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2629−2639
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9315

generalisasi karena penggunaan satu dataset, sehingga diperlukan penelitian lanjutan untuk validasi manual dan pengujian lintas dataset guna memastikan akurasi semantik sistem.

REFERENCES

[1] S. Li and W. Deng, "Deep Facial Expression Recognition: A Survey," *IEEE Trans. Affect. Comput.*, vol. 13, no. 3, pp. 1195–1215, 2022, doi: 10.1109/TAFFC.2020.2981446.
[2] H. Shin, B. Lee, B. Ku, and H. Ko, "Noisy label facial expression recognition via face-specific label distribution learning," *Image Vis. Comput.*, vol. 143, p. 104901, Mar. 2024, doi: 10.1016/j.imavis.2024.104901.
[3] K. Wang, X. Peng, J. Yang, S. Lu, and Y. Qiao, "Suppressing uncertainties for large-scale facial expression recognition," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 6896–6905, 2020, doi: 10.1109/CVPR42600.2020.00693.
[4] B. Jiang *et al.*, "Research on facial expression recognition algorithm based on improved MobileNetV3," *EURASIP J. Image Video Process.*, vol. 2024, no. 1, p. 22, Aug. 2024, doi: 10.1186/s13640-024-00638-z.
[5] I. D. Mienye, T. G. Swart, and G. Obaido, "Recurrent Neural Networks: A Comprehensive Review of Architectures, Variants, and Applications," *Information*, vol. 15, no. 9, p. 517, Aug. 2024, doi: 10.3390/info15090517.
[6] Z. Zhao, Q. Liu, and S. Wang, "Learning Deep Global Multi-Scale and Local Attention Features for Facial Expression Recognition in the Wild," *IEEE Transactions on Image Processing*, vol. 30, pp. 6544–6556, 2021, doi: 10.1109/TIP.2021.3093397.
[7] S. Minaee, M. Minaei, and A. Abdolrashidi, "Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network," *Sensors*, vol. 21, no. 9, p. 3046, Apr. 2021, doi: 10.3390/s21093046.
[8] S. Ullah, J. Ou, Y. Xie, and W. Tian, "Facial expression recognition (FER) survey: a vision, architectural elements, and future directions," *PeerJ Comput. Sci.*, vol. 10, p. e2024, Jun. 2024, doi: 10.7717/peerj-cs.2024.
[9] A. Khelifa, H. Ghazouani, and W. Barhoumi, "Label distribution learning for compound facial expression recognition in-the-wild: A comparative study," *Expert Syst.*, vol. 42, no. 2, Feb. 2025, doi: 10.1111/exsy.13724.
[10] Z. Shao *et al.*, "Facial Action Unit Detection by Adaptively Constraining Self-Attention and Causally Deconfounding Sample," *Int. J. Comput. Vis.*, vol. 133, no. 4, pp. 1711–1726, Apr. 2025, doi: 10.1007/s11263-024-02258-6.
[11] N. Begum and A. S. Mustafa, "CNN BLSTM Joint Technique on Dynamic Shape and Appearance of FACS," *Int. J. Eng. Adv. Technol.*, vol. 9, no. 4, pp. 1754–1757, 2020, doi: 10.35940/ijeat.d7308.049420.
[12] C. Liang and J. Dong, "A Survey of Deep Learning-based Facial Expression Recognition Research," *Frontiers in Computing and Intelligent Systems*, vol. 5, no. 2, pp. 56–60, 2023, doi: 10.54097/fcis.v5i2.12445.
[13] T. Kopalidis, V. Solachidis, N. Vretos, and P. Daras, "Advances in Facial Expression Recognition: A Survey of Methods, Benchmarks, Models, and Datasets," *Information (Switzerland)*, vol. 15, no. 3, 2024, doi: 10.3390/info15030135.
[14] D. Liang, H. Liang, Z. Yu, and Y. Zhang, "Deep convolutional BiLSTM fusion network for facial expression recognition," *Vis. Comput.*, vol. 36, no. 3, pp. 499–508, Mar. 2020, doi: 10.1007/s00371-019-01636-3.
[15] Y. Li, J. Zeng, and S. Shan, "Learning Representations for Facial Actions From Unlabeled Videos," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 1, pp. 302–317, Jan. 2022, doi: 10.1109/TPAMI.2020.3011063.
[16] S. Jayaraman and A. Mahendran, "An Improved Facial Expression Recognition using CNN-BiLSTM with Attention Mechanism," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 5, 2024, doi: 10.14569/IJACSA.2024.01505132.
[17] J. Zhong, T. Chen, and L. Yi, "Face expression recognition based on NGO-BILSTM model," *Front. Neurorobot.*, vol. 17, Mar. 2023, doi: 10.3389/fnbot.2023.1155038.
[18] B. H. Pansambal, A. B. Nandgaokar, J. L. Rajput, and A. Wagh, "An Integrated CNN-BiLSTM Approach for Facial Expressions," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 3, 2024, doi: 10.14569/IJACSA.2024.0150398.
[19] X. Ge, J. Fu, F. Chen, S. An, N. Sebe, and J. M. Jose, "Towards End-to-End Explainable Facial Action Unit Recognition via Vision-Language Joint Learning," in *Proceedings of the 32nd ACM International Conference on Multimedia*, New York, NY, USA: ACM, Oct. 2024, pp. 8189–8198. doi: 10.1145/3664647.3681443.
[20] I. D. Mienye and T. G. Swart, "A Comprehensive Review of Deep Learning: Architectures, Recent Advances, and Applications," *Information*, vol. 15, no. 12, p. 755, Nov. 2024, doi: 10.3390/info15120755.