

Analisis Performa Variasi Kedalaman ResNet pada Klasifikasi Emosi Berbasis Spektrogram EEG

¹*Sza Sza Amulya Larasati, ²Fitriyani
^{1,2}Telkom University
Bandung, Indonesia

¹ szaszaamulyalarasati@telkomuniversity.ac.id
² fitriyani@telkomuniversity.ac.id

*Penulis Korespondensi

Diajukan : 20/07/2026
Diterima : 25/07/2026
Dipublikasi : 30/07/2026

ABSTRAK

Electroencephalogram (EEG) merupakan sinyal fisiologis yang mampu merekam aktivitas listrik otak dan dapat dimanfaatkan dalam sistem pengenalan emosi. Karakteristik sinyal EEG yang kompleks, rawan *noise*, dan bersifat non-stasioner menyebabkan proses klasifikasi emosi memerlukan proses ekstraksi fitur dan kedalaman arsitektur model yang tepat agar mampu menangkap pola informasi yang relevan dalam mendeteksi emosi. Penelitian ini bertujuan menganalisis pengaruh variasi kedalaman arsitektur Residual Network (ResNet) menggunakan data EEG yang ditransformasikan menjadi spektrogram melalui proses Short-Time Fourier Transform (STFT). Sinyal disegmentasi setiap 30 detik dan setiap segmen menghasilkan sebuah spektrogram, sehingga setiap responden mampu menghasilkan lebih dari satu spektrogram pada stimulus yang sama. Citra spektrogram digunakan sebagai *input* untuk lima variasi arsitektur ResNet, yaitu ResNet-18, ResNet-34, ResNet-50, ResNet-101, dan ResNet-152. Model dilatih menggunakan dataset GAMEEMO yang merupakan dataset publik dengan stimulus *computer game* untuk menciptakan emosi tertentu. Untuk menghindari terjadinya *data leakage* karena segmen pada responden yang sama bisa tercampur ke data latih dan uji, evaluasi dilakukan dengan membagi data latih dan uji berbasis responden dengan nilai rasio kurang lebih 80:20. Hasil penelitian menunjukkan akurasi tertinggi diperoleh ResNet-152 sebesar 0.7500, presisi 0.7529, recall 0.7500, dan F1-score 0.7507. ResNet-34 juga mendapatkan hasil yang tidak jauh berbeda, dengan akurasi 0.7480, presisi 0.7600, recall 0.7480, dan F1-score 0.7505. Dengan selisih perbedaan akurasi yang hanya 0.002, ResNet-34 dinilai lebih efisien terutama dari segi komputasi karena memiliki jumlah parameter yang lebih sedikit, namun apabila berorientasi pada keakuratan prediksi memang ResNet-152 lebih unggul. Selain itu terjadi tren kenaikan dan penurunan akurasi pada lima kedalaman ResNet. Akurasi meningkat dari ResNet-18 ke ResNet-34, namun mengalami penurunan pada ResNet-50 kemudian mengalami peningkatan kembali pada ResNet-101 dan ResNet-152. Penelitian ini menunjukkan bahwa peningkatan kedalaman arsitektur ResNet tidak selalu menghasilkan peningkatan performa secara linier, meskipun ResNet-152 memberikan hasil klasifikasi terbaik.

Kata Kunci: EEG, kedalaman jaringan, klasifikasi emosi, ResNet, STFT, spektrogram

I. PENDAHULUAN

Emosi menjadi salah satu komponen penting yang memengaruhi berbagai aspek kehidupan manusia. Dalam model dimensi emosi, emosi direpresentasikan berdasarkan dua dimensi utama menjadi *valence* yang menggambarkan tingkat kepositifan atau kenegatifan emosi dan *arousal* yang menggambarkan tingkat aktivasi emosi (Alakus et al., 2020). Keadaan emosional tidak hanya menentukan bagaimana seseorang merespons suatu situasi, tetapi juga berperan dalam proses kognitif seperti pengambilan keputusan, pembentukan memori, persepsi, kemampuan

berpikir, serta interaksi sosial (Alarcão & Fonseca, 2019)(Arzu et al., 2026). Kondisi emosi juga berkaitan erat dengan kesehatan mental sehingga menjadi indikator penting dalam memahami keadaan psikologis (Lv et al., 2026). Seiring pesatnya perkembangan teknologi, kemampuan sistem cerdas dalam mengenali dan memahami emosi manusia mampu menciptakan interaksi yang lebih adaptif dan berorientasi pada kebutuhan pengguna (Dangi et al., 2026). Hal tersebut mendorong berkembangnya bidang Affective Computing dan Human-Centered Artificial Intelligence dalam mengembangkan sistem cerdas agar dapat mengenali serta menyesuaikan respons berdasarkan emosi pengguna. Salah satu pendekatan yang banyak dikembangkan dalam bidang tersebut adalah deteksi emosi dengan memanfaatkan sinyal Electroencephalography (EEG) (Kukhilava et al., 2025). Dibandingkan indikator eksternal lainnya, sinyal EEG mampu merekam aktivitas listrik otak secara langsung sehingga dinilai lebih objektif dalam merepresentasikan kondisi emosional (Nayem et al., 2026). EEG memiliki potensi besar untuk diterapkan pada berbagai domain, misalnya dalam mendeteksi emosi (Alakus et al., 2020) (Hamzah & Abdalla, 2024) (Q. Li et al., 2021a) (Liu et al., 2025), tingkat stres manusia (Larasati et al., 2024), (Larasati et al., 2026) (Rai & Singh, 2025) (Mane & Shinde, 2023), hingga memahami keadaan mental (Parsa et al., 2023) (Jan et al., 2025) dan kesehatan manusia (Darankoum et al., 2026).

Meskipun deteksi berbasis EEG telah banyak diimplementasikan di berbagai domain, proses klasifikasi masih menghadapi berbagai tantangan. Sinyal EEG rentan terhadap *noise*, interferensi dari lingkungan, serta artefak fisiologis terutama akibat kedipan mata dan pergerakan kepala (Acharya et al., 2025). Hal tersebut mendorong diperlukannya tahap *preprocessing* dan ekstraksi fitur yang tepat untuk menghilangkan *noise* tanpa menghilangkan informasi penting dari sinyal EEG. Penelitian sebelumnya telah memanfaatkan berbagai metode ekstraksi fitur untuk merepresentasikan sinyal EEG, salah satunya adalah Fast Fourier Transform (FFT) yang mengubah sinyal dari domain waktu ke domain frekuensi untuk memudahkan proses interpretasi (M. Li & Chen, 2021) (Q. Li et al., 2021b) (Q. Li et al., 2021a). FFT memiliki keunggulan dalam segi komputasi, namun FFT mengasumsikan bahwa sinyal yang diolah bersifat stasioner sehingga frekuensi yang dihasilkan bersifat global tanpa mempertahankan informasi temporal kapan sinyal tersebut muncul (van Drongelen, 2007). Padahal, sinyal EEG merupakan sinyal non-stasioner dengan frekuensi terus berubah terhadap waktu akibat dinamika aktivitas otak (Wang et al., 2023). Beberapa penelitian menerapkan Short-Time Fourier Transform (STFT) (Larasati et al., 2026) (Atla & Sharma, 2023)(Shayeste & Asl, 2023) yang membagi sinyal menjadi sejumlah segmen dalam *window* tertentu kemudian menerapkan FFT pada setiap segmen tersebut. STFT dapat menciptakan citra spektrogram yang menginterpretasikan distribusi energi sinyal terhadap waktu dan frekuensi secara simultan (Jiang et al., 2025). Spektrogram mampu mengatasi keterbatasan FFT yang menganggap sinyal bersifat stasioner sehingga aktivitas otak dapat ditangkap lebih komprehensif (van Drongelen, 2007).

Dalam melakukan tugas klasifikasi, arsitektur *deep learning* yakni Convolutional Neural Network (CNN) banyak digunakan karena kemampuannya dalam mengekstraksi fitur secara otomatis melalui serangkaian lapisan konvolusi (Badr et al., 2024). Citra spektrogram mampu merepresentasikan informasi pada domain waktu dan frekuensi sehingga mampu mempertahankan karakteristik sinyal EEG yang berkaitan dengan kondisi emosional dan digunakan sebagai *input* CNN (Muktadi et al., 2025). ResNet sebagai pengembangan CNN telah banyak digunakan dalam klasifikasi di berbagai domain. Keandalan ResNet didukung dengan adanya mekanisme *residual learning* yang mampu mengatasi masalah *vanishing gradient* pada pelatihan jaringan yang lebih dalam (He et al., 2015). *Residual learning* memanfaatkan mekanisme *skip connection* yang tetap menyediakan informasi pada *layer* sebelumnya di *layer* yang lebih dalam. Dalam mendeteksi emosi pada dataset SEED, ResNet-50 berhasil mencapai akurasi 0,9310 (Sidharth et al., 2023), sedangkan ResNet-18 memperoleh akurasi 0,8706 (Cheah et al., 2021). Pada domain serupa yakni dalam mendeteksi stres, ResNet-152 yang dikombinasikan dengan mekanisme Attention dilatih menggunakan spektrogram EEG memperoleh akurasi 0,9026 (Larasati et al., 2026). ResNet juga berhasil diimplementasikan dalam mendeteksi schizophrenia pada data EEG dan memperoleh akurasi 0,9923 (Siuly et al., 2023). Berbagai hasil penelitian tersebut menunjukkan bahwa arsitektur ResNet memiliki

kemampuan yang konsisten dalam mengekstraksi representasi fitur dari sinyal EEG.

Penelitian ini bertujuan untuk mengenali emosi manusia berdasarkan sinyal EEG dengan ekstraksi fitur menggunakan STFT untuk menghasilkan spektrogram. Data EEG yang digunakan adalah GAMEEMO yang mampu merepresentasikan emosi berdasarkan skala *valence* dan *arousal* (Alakus et al., 2020). Citra spektrogram selanjutnya digunakan sebagai *input* pada model ResNet di lima kedalaman yang berbeda untuk mengatasi gap penelitian sebelumnya yang hanya melakukan klasifikasi berbasis EEG pada salah satu kedalaman ResNet saja tanpa mengetahui pengaruh kedalaman jaringan terhadap performa klasifikasi. Melalui mekanisme *residual learning*, penelitian ini menganalisis pengaruh kedalaman jaringan terhadap performa klasifikasi untuk mengevaluasi *trade-off* antara efisiensi komputasi berdasarkan jumlah parameter di setiap jaringan dan akurasi prediksi.

II. STUDI LITERATUR

Penelitian Terdahulu

EEG merupakan sinyal fisiologis yang mampu merekam aktivitas listrik otak secara objektif sehingga banyak dimanfaatkan dalam berbagai penelitian analisis kondisi mental dan emosional. Namun, sinyal EEG memiliki karakteristik yang kompleks, bersifat nonlinier, serta memiliki amplitudo yang rendah sehingga rentan terhadap berbagai noise dan artefak. Kompleksitas data EEG membutuhkan proses ekstraksi fitur dan arsitektur yang tepat agar informasi dapat diinterpretasikan dengan tepat. Arsitektur yang terlalu dangkal menyebabkan informasi yang diperoleh selama proses ekstraksi fitur menjadi kurang efisien, namun arsitektur dengan jaringan yang sangat dalam mampu meningkatkan terjadinya overfitting pada informasi yang sedang dipelajari (He et al., 2015). Dalam mengatasi permasalahan tersebut, ResNet hadir dengan mengimplementasikan mekanisme skip connection yang mampu menyimpan informasi dari lapisan sebelumnya kemudian diteruskan ke lapisan yang lebih dalam. Mekanisme ini membantu menjaga aliran informasi dan gradien selama proses pelatihan, sehingga mampu mengurangi permasalahan vanishing gradient serta degradasi performa pada jaringan yang semakin dalam.

Beberapa penelitian telah dikaji menggunakan data EEG dalam mendeteksi berbagai tugas klasifikasi yang disajikan pada Tabel 1. Penelitian sebelumnya berhasil mengembangkan metode deteksi emosi berbasis EEG menggunakan machine learning sebagai *classifier* yakni SVM dengan dataset GAMEEMO. Melalui ekstraksi fitur DWT, akurasi yang dihasilkan adalah 0.5500 (Alakus et al., 2020). Pendekatan deep learning juga berhasil diimplementasikan menggunakan ResNet-50 dengan pendekatan transfer learning yang dilatih dengan data SEED. Sinyal EEG direpresentasikan dalam bentuk citra 62×62 melalui kombinasi fitur Mean Phase Coherence (MPC), Magnitude Squared Coherence (MSC), dan Differential Entropy (DE) sebelum diklasifikasikan oleh model. Pendekatan tersebut memperoleh akurasi 0,9310 pada skenario subject-dependent dan 0,7160 pada skenario subject-independent (Sidharth et al., 2023). Pada dataset yang sama, ResNet-18 berhasil diimplementasikan dan memperoleh akurasi 0,8706. Setelah dilakukan penyesuaian arsitektur melalui modifikasi dimensi convolutional kernel, akurasi model meningkat menjadi 0,9342 disertai dengan pengurangan jumlah parameter sebesar 52,22% (Cheah et al., 2021). STFT juga telah digunakan sebagai input model pada penelitian sebelumnya yang bertujuan untuk mengubah sinyal dalam domain *time-frequency* (Muktadi et al., 2025) (Larasati et al., 2026) (Larasati & Bachtiar, 2024). ResNet-152 dikombinasikan dengan Multihead Attention dalam mendeteksi tingkat stres manusia yang dilatih menggunakan data spektrogram EEG berhasil mendapatkan akurasi 0,9026. Sayangnya pembagian data pada penelitian tersebut tidak didasarkan pada subjek sehingga terdapat kemungkinan *data leakage*. Terlepas dari itu, representasi spektrogram hasil STFT yang dipadukan dengan arsitektur ResNet efektif dalam mengekstraksi karakteristik sinyal EEG yang kompleks sehingga mampu meningkatkan kinerja klasifikasi pada kondisi emosional berbasis EEG.

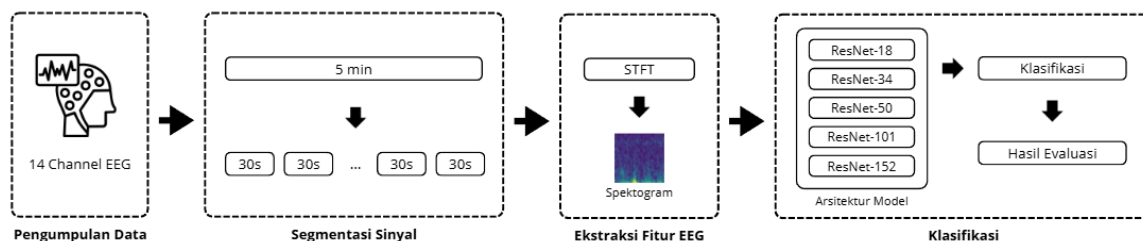
Meskipun pendekatan STFT dan ResNet berhasil diimplementasikan dalam beberapa domain kesehatan, penggunaan STFT sebagai metode ekstraksi fitur dalam melakukan klasifikasi emosi masih belum banyak diimplementasikan dibandingkan pendekatan ekstraksi fitur lainnya, padahal penggunaan STFT berpotensi menghasilkan performa klasifikasi yang lebih baik. Di sisi lain,

penelitian yang telah dikaji hanya mengevaluasi satu varian kedalaman arsitektur ResNet: ResNet-18, ResNet-50, dan ResNet-152 tanpa ada perbandingan evaluasi setiap kedalaman. Gap tersebut akan diselesaikan dalam penelitian ini menggunakan spektrogram yang didapatkan melalui STFT sebagai representasi sinyal EEG dan mengevaluasi lima variasi kedalaman arsitektur ResNet untuk mengidentifikasi model yang memberikan performa klasifikasi terbaik.

Tabel 1. Perbandingan hasil evaluasi penelitian sebelumnya

Referensi	Dataset	Ekstraksi Fitur	Arsitektur Model	Akurasi
(Sidharth et al., 2023)	SEED	MPC, MSC, DE	ResNet-50	0.9310
(Cheah et al., 2021)	SEED	Raw EEG	ResNet-18	0.8706
(Larasati et al., 2026)	Primer	STFT	ResNet-152+Attention	0.9026
(Alakus et al., 2020)	GAMEEMO	DWT	SVM	0.5500

III. METODE



Gambar 1. Graphical Abstract

Terdapat 4 tahapan utama dalam penelitian ini: pengumpulan data, segmentasi sinyal, ekstraksi fitur EEG, dan klasifikasi. Data yang digunakan merupakan dataset sekunder yakni GAMEEMO (Alakus et al., 2020). Selanjutnya sinyal disegmentasi untuk setiap 30 detik dan menerapkan STFT dalam setiap segmen tersebut. Hasil ekstraksi fitur berupa spektrogram selanjutnya digunakan sebagai input dalam model deep learning untuk melakukan klasifikasi emosi. Detail proses setiap tahapan dijelaskan rinci pada Section III ini.

Dataset

Data yang digunakan pada penelitian ini merupakan data sekunder yang dapat diakses publik yakni GAMEEMO. Dataset ini merupakan EEG yang direkam pada 28 responden berusia 20–27 tahun. Proses akuisisi data dilakukan dengan perangkat EMOTIV EPOC+, yaitu perangkat EEG portabel dengan 14 kanal (AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, dan AF4), sedangkan elektroda P3 dan P4 digunakan sebagai elektroda referensi. Stimulus yang diberikan selama perekaman adalah memainkan *computer game*. Sinyal EEG direkam pada rentang frekuensi 0,16–43 Hz dengan menerapkan filter sinc orde lima bawaan perangkat untuk mereduksi artefak akibat pergerakan kepala, tangan, dan gangguan eksternal lainnya. Dataset ini memiliki 4 label: *boring* (1), *calm* (2), *horror* (3), dan *funny* (4). Label tersebut ditentukan berdasarkan stimulus *computer game* yang dimainkan responden selama perekaman data (Alakus et al., 2020).

Segmentasi Sinyal

Sinyal EEG pada setiap responden yang semula berdurasi 5 menit disegmentasi menjadi beberapa bagian berdurasi 30 detik sebelum dilakukan ekstraksi fitur tanpa adanya *overlap*. Segmentasi dilakukan karena sinyal EEG bersifat nonstasioner, sehingga karakteristik sinyal dapat berubah seiring waktu. Durasi yang lebih singkat mampu menangkap perubahan spektral pada waktu lebih detail (Hendrawan et al., 2022). Selain itu, proses segmentasi juga meningkatkan jumlah sampel untuk pelatihan model sehingga mampu memperkaya variasi data dan membantu meningkatkan kemampuan generalisasi model dalam proses klasifikasi. Data yang digunakan memiliki 4 label

dengan 28 responden, sehingga total data sebelum dilakukan segmentasi adalah 112 dan setelah segmentasi menjadi 1120 data.

Ekstraksi Fitur EEG

Ekstraksi fitur memanfaatkan metode STFT yang menghasilkan citra spektrogram. STFT dirumuskan pada Persamaan (1). STFT memecah sinyal asli $x[n]$ menjadi segmen-segmen kecil atau yang biasa disebut dengan *window* sebanyak $x_m[n]$. Dengan ukuran *window* $nperseg$, sinyal akan terbagi dalam rentang $n \in [m, m + nperseg - 1]$, m adalah titik awal *window* tersebut. Pembagian ini memungkinkan analisis sinyal dalam potongan waktu yang lebih kecil, sehingga perubahan frekuensi dapat diproses secara lokal sepanjang waktu. Pada penelitian ini STFT dilakukan dengan *Hann window* pada panjang *window* 128 sampel, *overlap* 64 sampel (50%), dan panjang FFT 128 titik. Spektrogram dihasilkan pada setiap *channel* sehingga pada segmen yang sama dihasilkan 14 spektrogram.

$$x_m[n] = x[n] \text{ untuk } n \in [m, m + nperseg - 1] \quad (1)$$

Klasifikasi

Setiap citra yang dihasilkan dari setiap *channel* diperlakukan sebagai satu sampel independen dan setiap sampel menginterpretasikan label emosi terkait, sehingga jumlah data yang dihasilkan meningkat secara proporsional terhadap jumlah *channel*. Setelah melalui proses segmentasi dan ekstraksi fitur, jumlah data seimbang untuk setiap kelas dan detail jumlah data awal dan data setelah segmentasi ditunjukkan pada Tabel 2. Proses klasifikasi mengimplementasikan arsitektur *pretrained* ResNet pada lima kedalaman yang berbeda, yaitu ResNet-18, ResNet-34, ResNet-50, ResNet-101, dan ResNet-152. Operasi konvolusi pada ResNet sama halnya dengan CNN. ResNet mengadaptasi penggunaan blok residual melalui koneksi shortcut. Mekanisme ini akan melewati beberapa layer dengan melakukan identity mapping atau menambahkan *input* asli langsung ke *output* hasil pemrosesan lapisan konvolusi. Secara matematis, output dari blok residual dirumuskan dalam Persamaan (2). Perbedaan utama setiap kedalaman terletak pada struktur blok residual dan jumlah pengulangan pada setiap blok. ResNet-18 dan ResNet-34 menggunakan Basic Residual Block yang terdiri dua lapisan konvolusi pada setiap blok, sehingga shortcut connection langsung menghubungkan input ke *output* setelah dua lapisan konvolusi, sedangkan ResNet-50, ResNet-101, dan ResNet-152 menggunakan Bottleneck Residual Block terdiri dari tiga lapisan konvolusi pada setiap blok. Detail jumlah layer dan blok residual untuk setiap kedalaman ditunjukkan pada Gambar 2. Evaluasi terhadap setiap varian dilakukan untuk mengetahui pengaruh kedalaman jaringan terhadap kemampuan ekstraksi fitur, akurasi klasifikasi, serta efektivitas model dalam mengenali emosi berdasarkan citra spektrogram EEG. Ukuran citra yang digunakan sebagai input diubah menjadi 256×256 , kemudian dilakukan dipangkas menjadi 224×224 . Proses klasifikasi dilakukan pada 200 epoch dengan mekanisme *early stopping* pada *patience* 20. *Hyperparameter* yang digunakan ditunjukkan pada Tabel 3.

$$Y = F(X) + X \quad (2)$$

Tabel 2. Detail jumlah data sebelum dan sesudah segmentasi-ekstraksi fitur

Data	Total	Data train setiap kelas	Data test setiap kelas
Awal	112 file	-	-
Setelah segmentasi dan ekstraksi fitur	14112 gambar	2772	756

Tabel 3. *Hyperparameter* ResNet yang digunakan

<i>Hyperparameter</i>	Nilai
<i>Optimizer</i>	SGD
<i>Batch size</i>	32
<i>Num workers</i>	4
<i>Learning rate</i>	0.001
<i>Momentum</i>	0.9

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
		3×3 max pool, stride 2				
conv2_x	56×56	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
conv4_x	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

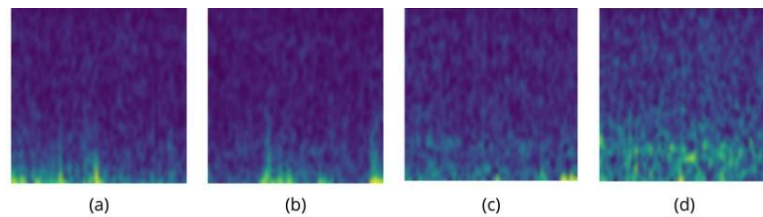
Gambar 2. Jumlah layer pada setiap kedalaman ResNet

Evaluasi

Untuk menghindari terjadinya *data leakage*, data uji dan data latih dibagi berdasarkan responden alih-alih spektrogram yang dihasilkan. Data yang dihasilkan pada responden 1 - 6 digunakan sebagai data uji, sedangkan responden 7 - 28 sebagai data latih. Hasil klasifikasi ResNet di setiap kedalaman selanjutnya dievaluasi menggunakan metrik akurasi, presisi, recall, dan F1-score untuk mengukur kinerja model secara menyeluruh. Akurasi mengukur proporsi sampel yang berhasil diklasifikasikan dengan benar terhadap seluruh data uji. Presisi menunjukkan ketepatan model dalam memberikan prediksi pada suatu kelas, sedangkan recall mengukur kemampuan model dalam mengidentifikasi seluruh sampel yang benar-benar termasuk ke dalam kelas tersebut. F1-score merupakan rata-rata harmonik antara presisi dan recall yang memberikan ukuran kinerja klasifikasi secara lebih seimbang.

IV. HASIL DAN PEMBAHASAN

Proses ekstraksi menghasilkan citra spektrogram untuk setiap channel EEG pada setiap segmen sinyal, sehingga setiap responden menghasilkan 14 channel × 10 segmen citra spektrogram untuk setiap kelas emosi. Sampel gambar di setiap label pada channel, segmen, dan responden yang sama ditunjukkan pada Gambar 3. Secara visual, keempat spektrogram meskipun digenerate berdasarkan segmen waktu dan channel yang sama, tampak perbedaan pada intensitas warna dan penyebaran energi di beberapa bagian spektrogram. Kelas boring (Gambar 3a) dan calm (Gambar 3b) menunjukkan pola spektrogram yang relatif serupa dengan dominasi warna ungu serta memiliki kecenderungan pada intensitas yang lebih rendah. Pada kelas horror (Gambar 3c), terlihat peningkatan penyebaran area berwarna hijau, sedangkan kelas funny (Gambar 3d) memperlihatkan distribusi warna hijau yang lebih dominan.



Gambar 3. Sampel data spektrogram yang dihasilkan pada segmen 4 (detik 201-230) pada channel AF4 untuk kelas (a) *boring*, (b) *calm*, (c) *horror*, dan (d) *funny*

Jumlah parameter pada masing-masing kedalaman ResNet ditunjukkan pada Tabel 4. Peningkatan jumlah *layer* berbanding lurus dengan jumlah parameter yang digunakan sehingga kompleksitas dan kebutuhan komputasi model turut meningkat. Hasil evaluasi untuk setiap kelas pada lima kedalaman ResNet ditunjukkan pada Tabel 5. ResNet-18, ResNet-101, dan ResNet-152 memperoleh akurasi tertinggi di kelas 1 yakni sebesar 0.8124, 0.8995, dan 0.8042, namun akurasi pada kelas 2 menghasilkan performa terendah 0.5040, 0.5979, 0.6892. ResNet-34 dan ResNet-50 juga menunjukkan pola serupa dengan akurasi tertinggi di kelas 1 yakni sebesar 0.7857 dan 0.7738, namun mendapatkan akurasi terendah pada kelas 4 yakni sebesar 0.6878 dan 0.6548. Hasil tersebut menunjukkan adanya *trade-off* antar kelas, model mampu mengenali kelas 1 dengan cukup baik tetapi kurang handal dalam mengenali kelas 2 dan 5. Meskipun model dengan *layer* dan parameter yang lebih banyak memiliki kapasitas representasi yang lebih besar, peningkatan jumlah *layer* tidak selalu menghasilkan peningkatan performa yang konsisten pada seluruh kelas dan hanya memberikan keuntungan pada kelas tertentu, seperti kelas 1.

Metrik akurasi, presisi, recall, dan F1-score untuk seluruh kelas ditunjukkan pada Tabel 5. ResNet-18 menghasilkan akurasi terendah sebesar 0.6852, kemudian meningkat pada ResNet-34 menjadi 0.7480. Namun, akurasi mengalami penurunan pada ResNet-50 menjadi 0.7136 sebelum kembali meningkat pada ResNet-101 dan mencapai nilai tertinggi pada ResNet-152 sebesar 0.7500. Perbedaan akurasi antara ResNet-34 dan ResNet-152 relatif kecil, yaitu hanya sebesar 0.0020. Hasil ini menunjukkan bahwa penambahan kedalaman jaringan tidak selalu memberikan peningkatan performa tetapi dipengaruhi pula oleh karakteristik arsitektur yang digunakan. ResNet-18 dan ResNet-34 menggunakan Basic Block, sedangkan ResNet-50, ResNet-101, dan ResNet-152 menggunakan Bottleneck Block yang memiliki kapasitas representasi fitur lebih tinggi. Meskipun demikian, peningkatan kapasitas model belum secara langsung menghasilkan peningkatan akurasi pada setiap varian.

Tabel 5 juga menunjukkan nilai presisi ResNet-34 bernilai lebih tinggi jika dibandingkan dengan 4 model lainnya. Nilai *precision* yang tinggi mengindikasikan model mampu mengurangi kesalahan prediksi, sehingga sampel yang diprediksi sebagai suatu kelas cenderung benar. Namun, nilai *recall* ResNet-34 masih sedikit lebih rendah dibandingkan ResNet-152. Hal ini menunjukkan bahwa meskipun ResNet-34 lebih selektif dalam memberikan prediksi, masih terdapat beberapa sampel yang seharusnya teridentifikasi tetapi belum berhasil dikenali. Meskipun keduanya berbeda keunggulan dalam metrik tertentu, perbedaan yang dihasilkan tidak besar sehingga F1-score yang merupakan rata-rata harmonik presisi dan *recall* juga memiliki nilai yang sangat mirip, yakni 0.7505 untuk ResNet-34 dan 0.7507 untuk ResNet-152.

Perbandingan nilai metrik baik akurasi, presisi, *recall*, dan F1-score pada ResNet-34 dan ResNet-152 tidaklah berbeda jauh. Jika dibandingkan berdasarkan jumlah parameter, ResNet-34 memiliki parameter yang jauh lebih sedikit dibandingkan dengan ResNet-152. Perbedaan jumlah *layer* keduanya hanya berdampak pada selisih akurasi sebesar 0.0020 pada penelitian ini. ResNet-34 dapat menjadi alternatif apabila berorientasi pada kompleksitas model dan kebutuhan komputasi, sedangkan ResNet-152 lebih unggul apabila tujuan utama adalah memperoleh performa klasifikasi yang paling seimbang.

Tabel 4. Jumlah parameter pada setiap kedalaman ResNet

Model	Jumlah Parameter (juta)
ResNet-18	11.689512
ResNet-34	21.797672
ResNet-50	25.557032
ResNet-101	44.549160
ResNet-152	60.192808

Tabel 5. Akurasi perlabel di setiap kedalaman ResNet

Model	Kelas			
	1	2	3	4
ResNet-18	0.8124	0.5040	0.7368	0.6786
ResNet-34	0.7857	0.7593	0.7593	0.6878
ResNet-50	0.7738	0.7090	0.7169	0.6548
ResNet-101	0.8995	0.5979	0.7513	0.7288
ResNet-152	0.8042	0.6892	0.7526	0.7540

Tabel 6. Hasil klasifikasi pada setiap kedalaman ResNet

Model	Akurasi	Presisi	Recall	F1-score
ResNet-18	0.6852	0.6896	0.6852	0.6817
ResNet-34	0.7480	0.7600	0.7480	0.7505
ResNet-50	0.7136	0.7211	0.7136	0.7151
ResNet-101	0.7444	0.7525	0.7444	0.7411
ResNet-152	0.7500	0.7529	0.7500	0.7507

Penurunan performa pada ResNet-50 menunjukkan bahwa peningkatan kedalaman model tidak selalu diikuti oleh peningkatan akurasi klasifikasi. Meskipun ResNet-50 memiliki kapasitas representasi yang lebih besar jika dibandingkan ResNet-34, jumlah parameter yang lebih banyak memerlukan data pelatihan yang lebih beragam agar dapat dipelajari secara optimal. Apabila data yang dipelajari kurang beragam, bertambahnya jumlah parameter dapat meningkatkan risiko *overfitting*, yaitu kondisi ketika model terlalu memahami pola pada data latih sehingga kemampuan generalisasi menurun pada data uji. Dataset yang digunakan terbatas pada 28 responden dengan pelabelan emosi berdasarkan stimulus meskipun sudah mempertimbangkan skala *valence* dan *arousal*.

Dengan menggunakan dataset yang sama, hasil penelitian ini menunjukkan bahwa kombinasi ekstraksi fitur menggunakan STFT dan klasifikasi menggunakan ResNet-152 mampu mengungguli penelitian sebelumnya yang ditunjukkan pada Tabel 7. Dalam mendeteksi emosi pada 4 kelas yang berbeda, penelitian ini memperoleh akurasi sebesar 0.7500, sedangkan penelitian terdahulu hanya mencapai akurasi sebesar 0.5500 dengan menggunakan ekstraksi fitur berbasis *Discrete Wavelet Transform* (DWT) dan model *Support Vector Machine* (SVM). Representasi *time-frequency* yang dihasilkan melalui STFT mampu mempertahankan informasi temporal dan spektral secara simultan, sehingga menyediakan informasi yang lebih representatif bagi proses pembelajaran. Selain itu, ResNet-152 yang memiliki ekstraksi fitur melalui konvolusi mampu melakukan ekstraksi fitur secara hierarkis lebih efektif dalam mengenali pola emosi pada sinyal EEG. Peningkatan akurasi tersebut tidak dapat sepenuhnya diatribusikan pada penggunaan STFT atau ResNet-152 secara terpisah, melainkan merupakan hasil dari kombinasi metode ekstraksi fitur dan model klasifikasi yang digunakan.

Tabel 7. Perbandingan akurasi dengan penelitian lainnya

Reference	Ekstraksi Fitur	Model	Akurasi
(Alakus et al., 2020)	DWT	KNN	0.4286
		SVM	0.5500
Penelitian ini	STFT	ResNet-152	0.7500

V. KESIMPULAN

Penelitian ini berhasil mengimplementasikan STFT dalam proses ekstraksi fitur EEG dan ResNet sebagai *classifier* dalam mendeteksi emosi dalam 4 jenis yang berbeda sesuai dengan skala *valence* dan *arousal*. Perbandingan klasifikasi dilakukan pada lima variasi kedalaman arsitektur ResNet, yaitu ResNet-18, ResNet-34, ResNet-50, ResNet-101, dan ResNet-152. Hasil pengujian berdasarkan konfigurasi penelitian ini dengan dataset GAMEEMO menunjukkan bahwa ResNet-152 memberikan performa terbaik dengan akurasi sebesar 0.7500 disusul oleh ResNet-34 yang memiliki akurasi 0.7480. Berdasarkan hasil pengujian, peningkatan kedalaman jaringan memberikan kecenderungan peningkatan performa klasifikasi, meskipun tidak bersifat linier. Dengan perbedaan performa yang relatif kecil, ResNet-34 menawarkan efisiensi komputasi yang lebih baik karena jumlah parameter yang lebih sedikit dibandingkan ResNet-152. Meskipun demikian, ResNet-152 tetap menghasilkan akurasi tertinggi karena kedalaman jaringan yang lebih besar meningkatkan kapasitas model dalam mengekstraksi fitur kompleks spektrogram EEG. Evaluasi model hanya dilakukan pada satu konfigurasi parameter STFT dan satu skenario pembagian data, sehingga hasil yang didapatkan bisa jadi belum memiliki kemampuan generalisasi yang sama jika diimplementasikan pada dataset EEG lainnya. Terlepas performa yang diperoleh, akurasi klasifikasi masih berpotensi ditingkatkan. Penelitian selanjutnya dapat melakukan eksplorasi pengaruh parameter STFT, seperti ukuran *window* dan tingkat *overlap* yang mampu memengaruhi representasi *time-frequency* yang dihasilkan. Selain itu, pengembangan arsitektur klasifikasi seperti optimasi *hyperparameter* dan penambahan mekanisme Attention juga berpotensi meningkatkan performa klasifikasi. Pembagian data juga dapat mempertimbangkan mekanisme *subject-wise cross-validation* untuk mengevaluasi kemampuan generalisasi model terhadap data dari subjek yang belum pernah digunakan selama proses pelatihan.

VI. REFERENSI

- Acharya, S., Khosravi, A., Creighton, D., Alizadehsani, R., & Acharya, U. R. (2025). Neurostressology: A systematic review of EEG-based automated mental stress perspectives. *Information Fusion*, 124, 103368. <https://doi.org/https://doi.org/10.1016/j.inffus.2025.103368>
- Alakus, T. B., Gonen, M., & Turkoglu, I. (2020). Database for an emotion recognition system based on EEG signals and various computer games – GAMEEMO. *Biomedical Signal Processing and Control*, 60, 101951. <https://doi.org/https://doi.org/10.1016/j.bspc.2020.101951>
- Alarcão, S. M., & Fonseca, M. J. (2019). Emotions Recognition Using EEG Signals: A Survey. *IEEE Transactions on Affective Computing*, 10(3), 374–393. <https://doi.org/10.1109/TAFFC.2017.2714671>
- Arzu, G. E., Umar, M., Khan, A., Ali, U., Dang, L. M., & Moon, H. (2026). Adaptive multimodal emotion detection for mental health monitoring using deep learning. *Information Sciences*, 744, 123385. <https://doi.org/https://doi.org/10.1016/j.ins.2026.123385>
- Atla, K. R., & Sharma, R. (2023). CNN based Human Intention-Behavior Prediction using Common Average Reference and STFT of EEG Signals. *2023 6th International Conference on Recent Trends in Advance Computing (ICRTAC)*, 65–70. <https://doi.org/10.1109/ICRTAC59277.2023.10480758>

- Badr, Y., Tariq, U., Al-Shargie, F., Babiloni, F., Al Mughairbi, F., & Al-Nashash, H. (2024). A review on evaluating mental stress by deep learning using EEG signals. *Neural Computing and Applications*, 36(21), 12629–12654. <https://doi.org/10.1007/s00521-024-09809-5>
- Cheah, K. H., Nisar, H., Yap, V. V., Lee, C.-Y., & Sinha, G. R. (2021). Optimizing Residual Networks and VGG for Classification of EEG Signals: Identifying Ideal Channels for Emotion Recognition. *Journal of Healthcare Engineering*, 2021(1), 5599615. <https://doi.org/https://doi.org/10.1155/2021/5599615>
- Dangi, D., Kumar, J., Gupta, R., & Sharma, S. (2026). Facial Emotion Detection using Optimized CNN and ResNet with PSO. *Procedia Computer Science*, 283, 835–843. <https://doi.org/https://doi.org/10.1016/j.procs.2026.06.165>
- Darankoum, D., Villalba, M., Allieux, C., Caraballo, B., Dumont, C., Gronlier, E., Roucard, C., Roche, Y., Habermacher, C., Grudin, S., & Volle, J. (2026). From epilepsy seizure classification to detection: A deep learning-based approach for raw EEG signals. *Neuroscience Informatics*, 6(1), 100263. <https://doi.org/https://doi.org/10.1016/j.neuri.2026.100263>
- Hamzah, H. A., & Abdalla, K. K. (2024). EEG-based emotion recognition systems; comprehensive study. *Heliyon*, 10(10), e31485. <https://doi.org/https://doi.org/10.1016/j.heliyon.2024.e31485>
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. *ArXiv*. <http://arxiv.org/abs/1512.03385>
- Hendrawan, M. A., Rosiani, U. D., & Sumari, A. D. W. (2022). Single Channel Electroencephalogram (EEG) Based Biometric System. *2022 IEEE 8th Information Technology International Seminar (ITIS)*, 307–311. <https://doi.org/10.1109/ITIS57155.2022.10010103>
- Jan, D., Rico, E., Birba, A., Ravelo, Y., León-Mendez, M., Travina, K., Padrón, I., de Vega, M., & Marrero, H. (2025). Non-Invasive Classification of Mental Health Disorders Using Resting-State EEG with Dry Electrodes for Scalable Triage. *Cognitive Computation*, 17(6), 171. <https://doi.org/10.1007/s12559-025-10529-8>
- Jiang, W., Gu, X., Zhao, Y., Tang, E., Jiang, X., Song, Z., & Zeng, L. (2025). Intelligent Fault Diagnosis of Hydraulic Pumps Based on Multi-Source Signal Fusion and Dual-Attention Convolutional Neural Networks. *Sensors*, 25(22), 7018. <https://doi.org/10.3390/s25227018>
- Kukhilava, N., Tsmindashvili, T., Kalandadze, R., Gupta, A., Katamadze, S., Brémond, F., Ferrari, L. M., Müller, P., & Wirth, B. E. (2025). *Evaluation in EEG Emotion Recognition: State-of-the-Art Review and Unified Framework*. <https://arxiv.org/abs/2505.18175>
- Larasati, S. S. A., & Bachtiar, F. A. (2024). Harnessing Residual Attention Networks for Stress Level Classification Using EEG Spectrograms. *2024 Ninth International Conference on Informatics and Computing (ICIC)*, 1–6. <https://doi.org/10.1109/ICIC64337.2024.10956441>
- Larasati, S. S. A., Bachtiar, F. A., & Setiawan, B. D. (2024). Leveraging EEG in Classifying Human Stress With Machine Learning Approach. *2024 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, 64–69. <https://doi.org/10.1109/IECBES61011.2024.10991282>
- Larasati, S. S. A., Bachtiar, F. A., & Setiawan, B. D. (2026). Multi-Band EEG Spectrogram Decomposition with Residual Attention Network for Enhanced Stress Classification. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 10(2), 325–331. <https://doi.org/10.29207/resti.v10i2.7225>
- Li, M., & Chen, W. (2021). FFT-based deep feature learning method for EEG classification. *Biomedical Signal Processing and Control*, 66, 102492. <https://doi.org/https://doi.org/10.1016/j.bspc.2021.102492>
- Li, Q., Liu, Y., Liu, C., Yan, F., Zhang, Q., Liu, Q., & Gao, W. (2021a). EEG signal processing and emotion recognition using Convolutional Neural Network. *2021 International Conference on Electronic Information Engineering and Computer Science (EIECS)*, 81–84. <https://doi.org/10.1109/EIECS53707.2021.9587900>
- Li, Q., Liu, Y., Liu, C., Yan, F., Zhang, Q., Liu, Q., & Gao, W. (2021b). EEG signal processing and emotion recognition using Convolutional Neural Network. *2021 International*

- Conference on Electronic Information Engineering and Computer Science (EIECS)*, 81–84. <https://doi.org/10.1109/EIECS53707.2021.9587900>
- Liu, J., Wu, G., Fu, Q., Luo, Y., Yang, S., Qiu, S., Cao, Y., & Li, W. (2025). EEG-based emotion detection using Long Short-Term Memory Network and reinforcement learning for enhanced feature selection. *Applied Soft Computing*, 182, 113512. <https://doi.org/https://doi.org/10.1016/j.asoc.2025.113512>
- Lv, B., Yu, Y., Zhuo, F., & Tang, F. (2026). Real-time EEG-based emotion recognition using Riemannian quantification learning. *Biomedical Signal Processing and Control*, 114, 109223. <https://doi.org/https://doi.org/10.1016/j.bspc.2025.109223>
- Mane, S. A. M., & Shinde, A. (2023). StressNet: Hybrid model of LSTM and CNN for stress detection from electroencephalogram signal (EEG). *Results in Control and Optimization*, 11, 100231. <https://doi.org/https://doi.org/10.1016/j.rico.2023.100231>
- Muktadi, Md. T., Tahmeed Reza, T., Hossain, M. T., Jahan Arifa, E., & Karim Khondakar, Md. F. (2025). Spectrogram-driven Emotion Detection from Electroencephalogram. 2025 *International Conference on Electrical, Computer and Communication Engineering (ECCE)*, 1–6. <https://doi.org/10.1109/ECCE64574.2025.11013815>
- Nayem, F. M., Prity, F. S., & Shiam, A. Al. (2026). Emotion detection using AI with explainable AI based perception: A comparative analysis between machine learning and deep learning. *Informatics and Health*, 3(2), 164–209. <https://doi.org/https://doi.org/10.1016/j.infoh.2026.06.004>
- Parsa, M., Rad, H. Y., Vaezi, H., Hossein-Zadeh, G.-A., Setarehdan, S. K., Rostami, R., Rostami, H., & Vahabie, A.-H. (2023). EEG-based classification of individuals with neuropsychiatric disorders using deep neural networks: A systematic review of current status and future directions. *Computer Methods and Programs in Biomedicine*, 240, 107683. <https://doi.org/https://doi.org/10.1016/j.cmpb.2023.107683>
- Rai, R. K., & Singh, D. K. (2025). Stress detection through wearable EEG technology: A signal-based approach. *Computers and Electrical Engineering*, 126, 110478. <https://doi.org/https://doi.org/10.1016/j.compeleceng.2025.110478>
- Shayeste, H., & Asl, B. M. (2023). Automatic seizure detection based on Gray Level Co-occurrence Matrix of STFT imaged-EEG. *Biomedical Signal Processing and Control*, 79, 104109. <https://doi.org/https://doi.org/10.1016/j.bspc.2022.104109>
- Sidharth, S., Samuel, A. A., H, R., Panachakel, J. T., & Parveen K, S. (2023). Emotion detection from EEG using transfer learning. 2023 *45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 1–4. <https://doi.org/10.1109/EMBC40787.2023.10340389>
- Siuly, S., Guo, Y., Alcin, O. F., Li, Y., Wen, P., & Wang, H. (2023). Exploring deep residual network based features for automatic schizophrenia detection from EEG. *Physical and Engineering Sciences in Medicine*, 46(2), 561–574. <https://doi.org/10.1007/s13246-023-01225-8>
- van Drongelen, W. (2007). 6 - Continuous, Discrete, and Fast Fourier Transform. In *Signal Processing for Neuroscientists* (pp. 91–105). Academic Press. <https://doi.org/https://doi.org/10.1016/B978-012370867-0/50006-1>
- Wang, C., Li, Y., Wang, L., Liu, S., & Yang, S. (2023). A study of EEG non-stationarity on inducing false memory in different emotional states. *Neuroscience Letters*, 809, 137306. <https://doi.org/https://doi.org/10.1016/j.neulet.2023.137306>